From cosmology to galaxy formation : what can we learn from the large-scale structure of the Universe ?

Sandrine Codis-Decara

Thèse de Doctorat de l'Université Pierre et Marie Curie

Spécialité: Astronomie & Astrophysique

Ecole Doctorale 127 "Astronomie et Astrophysique d'Île de France"
Institut d'Astrophysique de Paris

Presentée par

Sandrine Codis-Decara

Pour obtenir le grade de

Docteur de l'Université Pierre et Marie Curie

Sujet de la thèse:

De la cosmologie à la formation des galaxies : que nous apprennent les grandes structures de l'Univers?

soutenue le 15 septembre 2015 devant le jury composé de

M.	Christophe Pichon	directeur de thèse
M.	Dmitri Pogosyan	directeur de thèse
M.	Pedro Ferreira	rapporteur
M.	Roman Scoccimarro	rapporteur
M.	Nick Kaiser	examinateur
M.	Adi Nusser	examinateur
M.	Benoît Semelin	examinateur
M.	Francis Bernardeau	membre invité

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Christophe, for his creative and numerous ideas, his everyday involvement, constant encouragement and support. I will not forget the nice time spent working together, I discovered what research was all about and particularly enjoyed the analytical works carried out together and the insightful discussions in the RER about science and research!

Second, I would like to extend my dearest thanks to my co-supervisor, Dmitry, for his kindness and his theoretical expertise. I enjoyed and learnt a lot from our mathematical and cosmic web discussions! I would like also to warmly thank Dmitry and Lena for their hospitality during the two nice stays I spent in Edmonton.

I would like to thank my reporters Roman Scoccimarro and Pedro Ferreira for their patience given the size of this manuscript, together with the members of my thesis jury, Nick Kaiser, Adi Nusser and Benoit Semelin. I deeply appreciate their time and effort and I am grateful for their scientific interest in my work.

I would like also to express my dearest thanks to Francis Bernardeau as a director of IAP but on top of that for the works carried out together. I have learnt so much! His rigorousness, expertise and brilliant ideas are an incredible guidance and source of inspiration. Special thanks also go to Julien Devriendt who first makes me discover the beauty of the cosmic web and the other researchers with whom I have collaborated with great pleasure during the last four years, in particular Vincent Desjacques, Raphaël Gavazzi, Yohan Dubois, Adrianne Slyz and Elisa Chisari.

I want to thank the director of IAP, Laurent Vigroux and all the administrative and research staff for having provided me with ideal work conditions. Special thanks come in particular to Florence Durret and Valérie de Lapparent.

In addition, it is a pleasure to thank all my colleagues and friends at IAP with whom I shared nice discussions during lunchtime and coffee breaks: Thierry, Sébastien, Damien and all the PhD students in particular Alice and my office mates Vincent, Hayley and Charlotte.

Finally, I would like to thank my family for their unconditional love and support. Papa, maman, merci pour tout! Merci aussi à Florian et nos deux soleils Julie et Noélie, vous illuminez ma vie de vos sourires et votre amour!

De la cosmologie à la formation des galaxies: que nous apprennent les grandes structures de l'Univers ?

Résumé

Dans cette thèse sur articles, nous nous intéressons aux défis posés aux spécialistes des grandes structures de l'Univers par les futurs relevés de galaxies tels que la mission Euclid de l'Agence Spatiale Européenne ou encore le LSST. Ces grandes structures sont la manifestation de l'organisation spatiale des galaxies qui ne sont pas de simples îles peuplant aléatoirement notre Univers mais forment un réseau complexe, souvent appelé "toile cosmique", fait de filaments entourant de grands vides et s'intersectant pour former des amas de galaxies. L'étude des grandes structures de l'Univers joue un rôle fondamental pour répondre aux grandes questions posées par l'étude du cosmos:

- Quels sont les constituants de notre Univers et les lois qui régissent son évolution?
- Comment les galaxies se forment-elles et évoluent-elles? En particulier, quel est l'impact de l'environnement sur leur structuration?

En effet, les galaxies naissent et grandissent au sein de ces grandes autoroutes cosmiques soulevant la question de l'impact sur les propriétés galactiques telles que leur morphologie, de la dynamique des grands courants cosmiques desquels elles émergent. Pour étudier cette question fondamentale, nous allons dans un premier temps montrer que dans les simulations numériques de l'Univers, le spin des galaxies est fortement lié à la direction de leur filament hôte avec un comportement qui dépend de leur masse. Ces corrélations spin-filament seront expliquées qualitativement dans le contexte de la formation hiérarchique des structures cosmologiques qui voit se former en premier lieu les petits objets puis, par fusion des objets de plus en plus massifs. Un modèle analytique tenant compte de l'anisotropie de la toile cosmique complètera ce tableau en reproduisant les corrélations observées. Ces idées sont importantes pour comprendre la morphologie des galaxies mais aussi les alignements intrinsèques émanant de la cohérence de la toile cosmique et pouvant contaminer certaines sondes cosmologiques basées sur la mesure de l'astigmatisme cosmique. Nous allons en particulier mesurer cette contamination dans une simulation hydrodynamique de pointe qui permet la prise en compte de la dynamique non-linéaire du gaz, la physique baryonique et l'anisotropie des grandes structures.

D'autre part, la croissance des grandes structures est sensible à la fois au taux d'expansion de l'Univers et à la vitesse à laquelle la gravité produit l'effondrement gravitationnel, ce qui en fait une source importante d'information pour répondre aux mystères de la cosmologie telles que la nature de l'énergie noire. Dans la seconde partie de ce manuscrit, nous nous poserons la question de comment extraire efficacement de l'information de la toile cosmique en mesurant sa topologie et sa géométrie et en utilisant la théorie perturbative dans un régime quasi-linéaire, la pierre angulaire de ce travail reposant sur l'étude analytique de l'impact de l'effondrement non-linéaire des structures et des distorsions en espace des redshifts sur la statistique du champ de densité cosmique.

Mots Clés: cosmologie, grandes structures de l'Univers, théorie perturbative, lentillage gravitationnel faible, galaxies, halos de matiére noire.

From cosmology to galaxy formation : what can we learn from the large-scale structure of the Universe ?

Abstract

This thesis by publication is devoted to address some of the challenges that the field of large-scale structure studies needs to overcome to extract the marrow of the gigantic precision datasets that will be produced by future galaxy surveys like ESA's cornerstone Euclid mission and the Large Synoptic Survey Telescope. This large-scale structure arises as galaxies are not islands randomly distributed in the Universe but form a complex network – the so-called cosmic web – made of filaments that bound large voids and intersect at clusters of galaxies. The study of the large-scale structure of the Universe plays a paramount role in our quest to answer the fundamental questions at the heart of Cosmology:

- What are the constituents of our Universe and what laws dictate its behaviour?
- How do galaxies form and evolve? In particular, what is the importance of their environment in shaping them?

Indeed, the birth and evolution of galaxies occur within these large cosmic highways and the natural question which arises is whether the properties of galaxies, such as their morphology, retain a memory of the large-scale cosmic flows from which they emerge. To address this key question, we will first show that in cosmological simulations, the spin of galaxies and the direction of the filament they are embedded in are correlated and that these correlations are mass-dependent. This signal will be shown to be qualitatively understood in the context of hierarchical structure formation, where small galaxies form first and merge together to form larger ones. An analytic model which explicitly takes into account the anisotropy of the cosmic web will complement this qualitative understanding by reproducing the measured correlations. Those ideas are important to understand the evolution of galaxy morphology and also to understand the intrinsic alignments of galaxies arising from the large-scale coherence of the cosmic web that contaminate cosmological probes like cosmic shear experiments. We will in particular measure this contamination directly from a state-of-the-art hydrodynamical simulation that allows us to take into account non-linear gas dynamics, baryonic physics and the anisotropy of the cosmic web.

The large-scale structure growth is sensitive to both the cosmic expansion rate and the rate at which gravity pulls matter together. It therefore constitutes a powerful probe of the fundamental issues at the core of modern cosmology e.g. dark energy equation of state or modified gravity. In a second part, we will address the question of how to efficiently use large-scale structure data to probe the cosmological model describing our Universe by measuring its topology and geometry and perturbation theory in the weakly and even mildly non-linear regime. The major contribution of this work is to analytically study the effect of redshift space distortions and non-linear collapse of structures on the topology, geometry and statistics of the cosmic density field.

Keywords: cosmology, large-scale structure of the Universe, perturbation theory, weak gravitational lensing, galaxies, dark matter haloes.

Introduction

If, at first sight, astrophysical objects seem to be uniformly distributed in the sky, deeper investigations reveal the existence of clusters of stars and our own galaxy – the Milky Way – that crosses the sky. Astronomers have established the existence of numerous other galaxies that are not islands uniformly distributed in the sky but gather into clusters and superclusters of galaxies. Until the end of the previous century, those clusters were believed to be the largest structures in the Universe and to be organised in a random fashion. The cosmological principle, on which the standard model of our Universe relies, is built upon an assumption of isotropy and homogeneity which seems to be valid only on very large scales. Indeed, the eighties revealed that galaxies are not islands randomly distributed but form a bubble-like network on scales of hundreds of megaparsecs! The dynamics of matter on very large scales is now quite well established thanks to numerical simulations and large galaxy surveys. Matter escapes from the voids towards the walls then flows towards the filaments before accreting onto the over-dense nodes of this so-called cosmic web. This process takes place simultaneously at multiple scales and epochs.

Galaxy formation and evolution naturally take place within this large-scale cosmic dynamics, raising the fundamental question : what is the role of the cosmic web in shaping galaxies? Galactic masses are highly dependent on their large-scale surrounding as explained by the theory of biased clustering. High mass objects preferentially form in over-dense environments (dense filaments and nodes). Beyond their mass, how much are other galaxy properties affected by their environment? In particular how is the rotation axis of galaxies correlated with the large-scale filamentary structure of the Universe? These questions are crucial to understanding galaxy formation and are at the core of the so-called nature versus nurture competition.

The large-scale structure of the Universe is also a key probe of the cosmos. Indeed its evolution and growth across cosmic time is directly sensitive to the properties of the Universe; studying them may allow us to answer key questions of modern cosmology : What are the laws driving the Universe? Was Einstein right or can we measure deviations from the predictions of his theory of Gravitation namely General relativity? What are the constituents of the cosmos? What is the engine of cosmic acceleration?

For all the above-mentioned reasons, the large-scale structure of the Universe has become an active field of research in astronomy and is destined to be crucial for the future with upcoming large galaxy surveys like Euclid or LSST that will release an impressive amount of data about the large-scale Universe.

This thesis describes the research carried out during my PhD at the Institut d'Astrophysique de Paris with my advisor, Christophe Pichon, and co-advisor, Dmitry Pogosyan from September 2011 to September 2015. This work focused on a theoretical study of the large-scale structure of the Universe and its interplay with, on the one hand, cosmology and, on the other hand, galaxy formation. This manuscript will mainly present some developments done in trying to answer two fundamental questions:

- how can we efficiently use large-scale structure to probe cosmology?
- what is the role of the cosmic web in shaping galaxy, in particular regarding galactic spin acquisition?

More precisely, the manuscript is organised as follows. In the first part of the thesis, we introduce concepts which are important to understand the genesis of the large-scale structure and its interplay with cosmology and galaxy formation. In chapter 1, we describe the homogeneous model of our Universe. Chapter 2 is devoted to a theoretical description of the large-scale structure of the Universe and the so-called cosmic web as the result of the growth of primordial small inhomogeneities in the matter distribution under the laws of gravity. In chapter 3, we give an overview of how galaxies form and evolve with a particular emphasize on mass and spin acquisition and the role of the environment.

The second part of the thesis focuses on the results obtained during my PhD. Chapter 4 describes how the spin of simulated dark matter halos and galaxies is correlated with the large-scale filamentary cosmic web. Those correlations can be understood from first principles when the standard theory of spin acquisition by tidal torquing takes into account the anisotropic nature of the large-scale environment of galaxies. Eventually, the correlation between galaxy shapes and the cosmic web induces a possible contamination for weak lensing experiments (the so-called intrinsic alignments) that can be measured in simulations. Chapter 5 then presents the results obtained regarding the topology of the large-scale structure in redshift space and the cosmic density in concentric spheres, two promising probes of cosmology for the coming years. Finally, for the interested reader, appendix A gathers other publications of mine that I refer to in the main text but that I chose to put apart as my contribution was less significant.

Contents

Part I

Context

1 Cosmology

Our Universe is well described by a simple and successful model, often called Λ-Cold Dark Matter. In this chapter, we describe its ingredients, its successes and possible failures.

1.1 The concordant model

1.1.1 Thermal history

Recent years have seen the building of a concordant model of our Universe called Λ-CDM. In this model, the Universe emerges out of a Big-Bang singularity 13.8 billion years ago.The initial hot plasma then cools down as the space-time expands. Soon after the singularity ($t \lesssim 10^{-35}s$), a period of cosmic inflation may have made the Universe flat and homogeneous ([Guth, 1981]), followed by the electroweak phase transition around $t \lesssim 10^{-11}s$ that breaks the $SU(2) \times U(1)$ electroweak symmetry to the present-day $U(1)$ electromagnetic symmetry and by the quark-hadron transition ($t \lesssim 10^{-6}s$) that converts the quark-gluon plasma into hadrons. The Universe is now dominated by photons and enters the radiation-dominated era. These primordial phases of the cosmic thermal history are still very speculative because energies are very high (>GeV) but the end of the story is much more established, relying on a very well-known low-energy physics.

Big-Bang nucleosynthesis, between $t \sim 3$ and 20 minutes, permits the formation of light atomic nuclei like Hydrogen, Deuterium, Helium, Lithium and Beryllium. After 70,000 years, the density of photons and matter are equal and the Universe falls into a matter-dominated era during which dark matter perturbations start to grow. Recombination then occurs when neutral hydrogen can be created as electrons are captured by the atomic nuclei. At the end of recombination, photons decouple from matter and travel freely : the Universe becomes transparent. This first light emitted at $t \lesssim 380,000$ years is called the Cosmic Microwave Background (CMB) and has been observed and extensively studied since its discovery by Penzias and Wilson in 1965. It is one of the smoking guns of the Hot Big Bang scenario. The mean temperature of this almost perfect black body is cooling down as the Universe is expanding and is currently $\bar{T} \approx 2.725$ K for all directions in the sky up to tiny fluctuations of the order of 10^{-5}. This means that at recombination, the Universe is very homogeneous and isotropic as predicted by the cosmic inflation scenario. The deviations from homogeneity seen in the CMB are believed to originate from initial quantum fluctuations and to be the seeds that will form the large-scale structure.

After the dark ages, from 300 million to one billion years after the Big-Bang, the first stars and galaxies form and the reionization epoch begins. After nine billion years, the Universe starts to be dark energy dominated and therefore undergoes a late-time period of accelerated expansion. The large-scale structure continues to form on larger and larger scales and is another paramount piece of evidence of the cosmological model. This will be developed in the rest of this thesis.

1.1.2 Energy budget

The Λ-Cold Dark Matter (ΛCDM) scenario is based on a description of the Gravitation force by means of Einstein's theory of General Relativity. It requires the introduction of a component of dark energy through the constant Λ entering the Einstein equations in order to explain the acceleration of cosmic expansion and a component of cold dark matter to explain the growth of structure (and in particular the population of galaxies that we observe today). The energy content has been constrained to be approximatively 69% in dark energy, 26% in dark matter and 5% in baryons ([Planck Collaboration et al., 2014c]). Radiation only accounts for 10^{-5} of the energy budget today. The minimal number of parameters that describe this cosmological model is six : three for the energy content i.e physical baryon, dark matter and dark energy densities; two for the initial power spectrum of fluctuations : scalar spectral index n_s and curvature fluctuation amplitude A_s; and eventually the reionization optical depth τ. This model in particular assumes that the spatial curvature is zero.

If we believe in the concordant cosmological model, matter in the Universe is not dominated by the ordinary – baryonic – matter that we are familiar with on Earth but by a cold – i.e non-relativistic – component that does interact only through gravity and is called *cold dark matter*. This collisionless type of matter does not interact electromagnetically and therefore can not be directly detected. As it does not interact with radiation, dark matter fluctuations can start to grow as soon as the Universe becomes matter-dominated. Later, baryonic matter, once decoupled from photons, falls into the gravitational potential wells produced by the dark matter distribution which eventually leads to the formation of baryonic structures like galaxies. Therefore, without dark matter, galaxies would not have time to form. Dark matter candidates are numerous : weakly interactive massive particles like neutrinos or neutralinos, axions, Kaluza-Klein particles, etc, but no direct detection has been made so far. Nevertheless, indirect detections strongly agree on its existence: the first report was made by Zwicky in 1933 ([Zwicky, 1933]) who found that the dynamics of the Coma cluster could only be explained if it was 300 times more massive than what observations suggested, then rotation curves of galaxies also provided evidence for the presence of large dark matter halos surrounding them ([Rubin et al., 1980]), gravitational lensing and the CMB eventually put tight constraints on the amount of dark matter in the Universe.

If the matter content of our Universe is mysterious, its energy content is even more so. Indeed, it should be dominated today by a dark energy that acts as a repulsive force and accelerates the cosmic expansion. This component is revealed by the CMB which requires a flat space-time and therefore 70% of dark energy in addition to baryonic and dark matter, and by the evolution of supernova luminosities with redshift that can only be explained by the existence of dark energy. It has to be noted that this dark energy component was discovered only two decades ago by [Riess et al., 1998, Perlmutter et al., 1999]. Its nature is completely unknown : vacuum energy, dynamical scalar field, modification of gravity, the possibilities are numerous and question the pillars of our current cosmological model.

1.1.3 Successes and possible failures

This rather simple model is in agreement with plenty of observations especially on large scales. The power spectrum of density fluctuations that can be observed on a wide range of scales using the cosmic microwave background, galaxy surveys and cluster abundance is in very good agreement with the Λ-CDM scenario (see for instance figure 1.1). Wiggles in the power spectrum

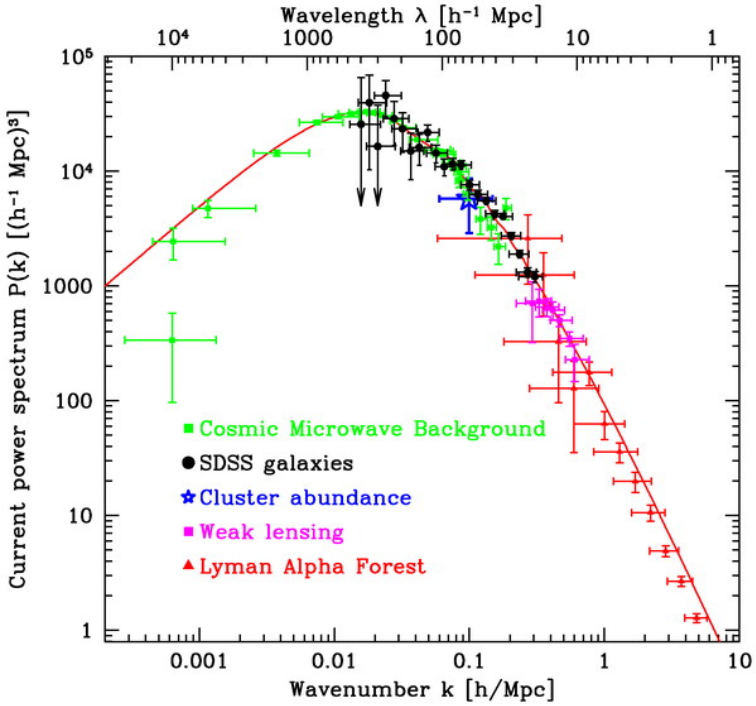


Figure 1.1: Galaxy power spectrum [Tegmark et al., 2004] measured on different scales using the CMB, cluster abundance, SDSS galaxies, weak lensing and the Lyman Alpha Forest as labelled. The solid line shows the Λ-CDM prediction for comparison.

named Baryon Acoustic Oscillations (BAOs) are also a great success of the concordant model [Blake et al., 2011b]. Those oscillations imprinted in the matter power spectrum come from the epoch of recombination when baryons were coupled to photons in a plasma and therefore were participating to acoustic waves propagating at the sound speed since the initial singularity. At recombination, baryons decouple and form structures preferentially at the location of the overdensities i.e separated by the sound horizon which is about 500 million light years. This feature in the matter power spectrum is tiny but the first oscillation has indeed been detected and constitute a standard ruler that can be measured at different epochs from the CMB (where several oscillations are visible as shown in figure 1.2) to the local Universe. Gravitational lensing and the Integrated Sachs-Wolfe effect also constitute strong evidence in favor of the concordant model.

Despite all the successes, on smaller scales where baryonic physics becomes important, the validity of this model is more difficult to probe. Three main problems have been emphasized as of today:

- the missing satellite problem : N-body simulations predict much more dark matter substructures than the observed number of galactic satellites;
- the angular momentum catastrophe : hydrodynamical simulations form disks too small compared to the observations;
- the cusp-core problem: dark matter simulations predict a central cuspy density profile for dark matter halos while observations find almost flat density profiles in the inner region of galaxies.

It is clear that the three above-mentioned problems are small-scale problems resulting from a

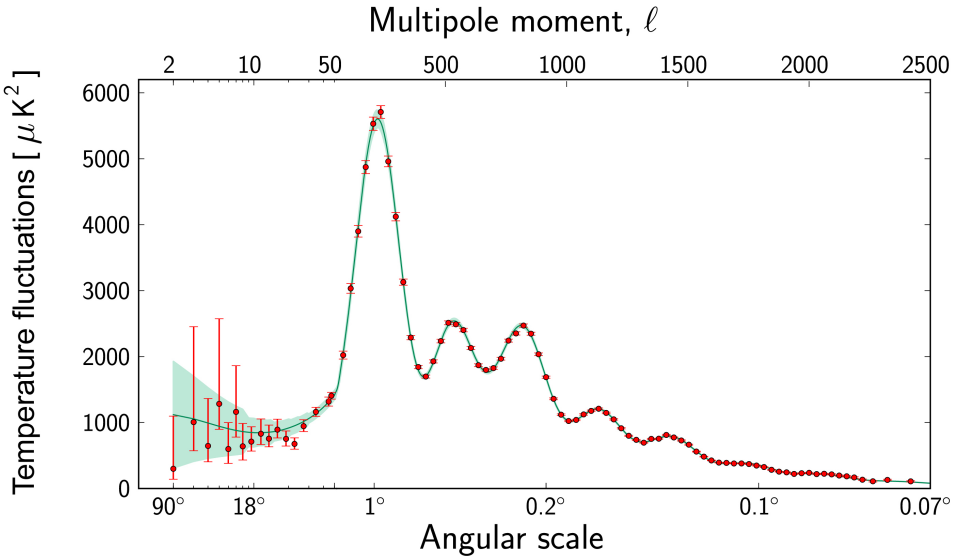


Figure 1.2: Temperature power spectrum of the CMB as measured by the Planck collaboration [Planck Collaboration et al., 2014b] and compared to the prediction of the best-fit Λ-CDM model.

comparison between the dark matter theory and observations of its baryonic counterpart. They can be resolved using either cosmological solutions or astrophysical arguments. Cosmological alternatives are to modify the small-scale power spectrum, introduce different types of dark matter (e.g warm dark matter that indeed removes small-scale fluctuations) or modify the Gravitation theory (e.g MOND that modifies General Relativity in the small acceleration regime). On the other hand, astrophysical solutions rely on upcoming improvements in our understanding of galaxy formation and argue that baryonic processes like stellar winds or feedback from active galactic nuclei may solve the problems. Galaxy formation issues will be developed in section 3.1.

Possible extensions of the Λ-CDM model include: modifications of gravity, other dark energy models probed through a parametrisation of the dark energy equation of state, inhomogeneous cosmological models, ...

1.2 The homogeneous Universe

In this section, we formalise the evolution of a homogeneous and isotropic universe which is one of the assumptions of the Λ-CDM model. The hypothesis of isotropy is well-established as the CMB and large galaxy surveys show statistical isotropy of the Universe on very large scales. The statistical homogeneity is then postulated assuming the Copernician Principle which states that we do not live in a special location of the Universe.

1.2.1 Hubble flow

A homogeneous and isotropic universe can be described using the Fridemann-Lemaitre-Robertson-Walker metric

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - R^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right] \tag{1.1}$$

where $R(t)$ is the scale factor ($a(t) = R(t)/R_0$ being its dimensionless equivalent) and $k = 0, \pm 1$ depends on the local curvature of space (flat, spherical or hyperbolic). The cosmic expansion rate is therefore characterised by the Hubble parameter $H = \dot{R}/R$. The expansion of the Universe, first discovered by Hubble in 1933, induces a red shift of the observed wavelengths compare to the emitted wavelengths that can be shown to be directly related to the scale factor

$$\frac{\lambda_{\mathrm{obs}}}{\lambda_{\mathrm{em}}} \equiv 1 + z = \frac{a(t_{\mathrm{obs}})}{a(t_{\mathrm{em}})}\,. \tag{1.2}$$

Radial distances are very difficult to measure. The best estimate we have is the redshift of the sources which, if they follow the Hubble flow i.e the cosmic expansion, is a direct probe of the distance. At small distance, the Hubble law that relates the redshift to the distance is recovered

$$z = \frac{H_0 d}{c}\,. \tag{1.3}$$

This assumption is not exactly true in practice as galaxies also have a peculiar velocity that affects its redshift (see also section 3.2.1).

1.2.2 Friedmann equation

Given the Fridemann-Lemaitre-Robertson-Walker metric, the dynamics of the Universe is then described by the evolution of $R(t)$ coming from the resolution of the Einstein equations of General Relativity

$$R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = 8\pi G T^{\mu\nu} + \Lambda g^{\mu\nu} \tag{1.4}$$

where the stress-energy tensor of a perfect fluid is written

$$T^{\mu\nu} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & -P \end{pmatrix}. \tag{1.5}$$

The resulting Friedmann equations then read

$$\begin{cases} \dot{R}^2 - \frac{8\pi G}{3}\rho R^2 = -kc^2 + \frac{\Lambda}{3}R^2, \\ \ddot{R} = -\frac{4}{3}\pi G R(\rho c^2 + 3P) + \frac{\Lambda}{3}R\,. \end{cases} \tag{1.6}$$

A simple conservation equation can be derived from the Friedmann equations (and is directly the result of the matter conservation $\nabla_\mu T^{\mu\nu} = 0$), $\rho R^{3(\omega+1)} =$ constant assuming an equation of state $P = \omega\rho$ with constant ω. Pressureless matter has $\omega = 0$, while for radiation $\omega = 1/3$. If Λ is interpreted as a cosmological fluid such that $\rho_\Lambda = -P_\Lambda/c^2 = \Lambda/(8\pi G)$, then it is described

by $\omega = -1$. Finally, the curvature can be interpreted as a fluid with equation of state $\omega = -1/3$ and $\rho_K = -3kc^2/(8\pi G R^2)$. Note that $k = 0$ corresponds to the concordance *Lambda*CDM model.

The critical density is

$$\rho_c = \frac{3H^2}{8\pi G}\,. \tag{1.7}$$

Its current value is $3H_0^2/8\pi G = 1.9\times10^{-29}h^2\,\mathrm{g/cm^3}$ where $H_0 = 100h$ km/s. Density parameters for matter, radiation and vacuum energy (Λ) can then be defined as $\Omega = \rho/\rho_c$ and the first Friedmann equation reads

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_K = 1\,. \tag{1.8}$$

With the conservation equation, it yields

$$\frac{H(z)}{H_0} = \sqrt{\Omega_{\Lambda\,0} + \Omega_{K\,0}(1+z)^2 + \Omega_{m\,0}(1+z)^3 + \Omega_{r\,0}(1+z)^4}\,. \tag{1.9}$$

1.2.3 Background expansion at different epochs

The matter-dominated era In the matter dominated era where radiation and relativistic matter do not contribute significantly to the energy budget, pressure can be neglected compared to the energy density ($\omega = 0$) leading to

$$\begin{cases} \rho(t)R^3(t) = \rho_0 R_0^3, \\ \dot{R}^2(t) = \frac{c^2}{|\Omega_{K\,0}|}\left[\Omega_{m\,0}\frac{R_0}{R} + \Omega_{\Lambda\,0}\left(\frac{R}{R_0}\right)^2 + \Omega_{K\,0}\right], \end{cases} \tag{1.10}$$

where the index 0 means that the quantities are taken at present time, $\Omega_K = kc^2/(H^2R^2)$ and $\Omega_\Lambda = \Lambda/(3H^2)$.

When the cosmological constant is zero, three cases appear. For $\Omega_{m\,0} = 1$, the Universe is flat and infinitely expanding according to $a(t) = (3/2H_0t)^{2/3}$. It is known as the Einstein-de Sitter Universe. For $\Omega_{m\,0} > 1$, the Universe is closed while it is open for $\Omega_{m\,0} < 1$.

When the cosmological constant is non-zero, the zoology of possible universes is broader and typically needs to be solved numerically.

The radiation-dominated era If our present-day Universe is matter-dominated, it does not hold anymore in our past as photons contribute more and more to the total energy density. Indeed, the matter density ρ_m is proportional to $(1+z)^3$ while the radiation density ρ_r goes like $(1+z)^4$ (owing to the volume dilution effect and the wavelength redshift). Therefore, when z increases, there necessarily exists a time when both components are equal. This equivalence between both component occurs at $z = z_{eq}$. Before this, the Universe is radiation-dominated, with $R(t) \propto t^{1/2}$, meaning that the Universe starts from an initial singularity.

The Λ-dominated era If the cosmological constant dominates, for zero curvature, the scale factor grows exponentially with time $R(t) \propto e^{Ht}$ where H is constant. It is known as the de Sitter space and the expansion is accelerated.

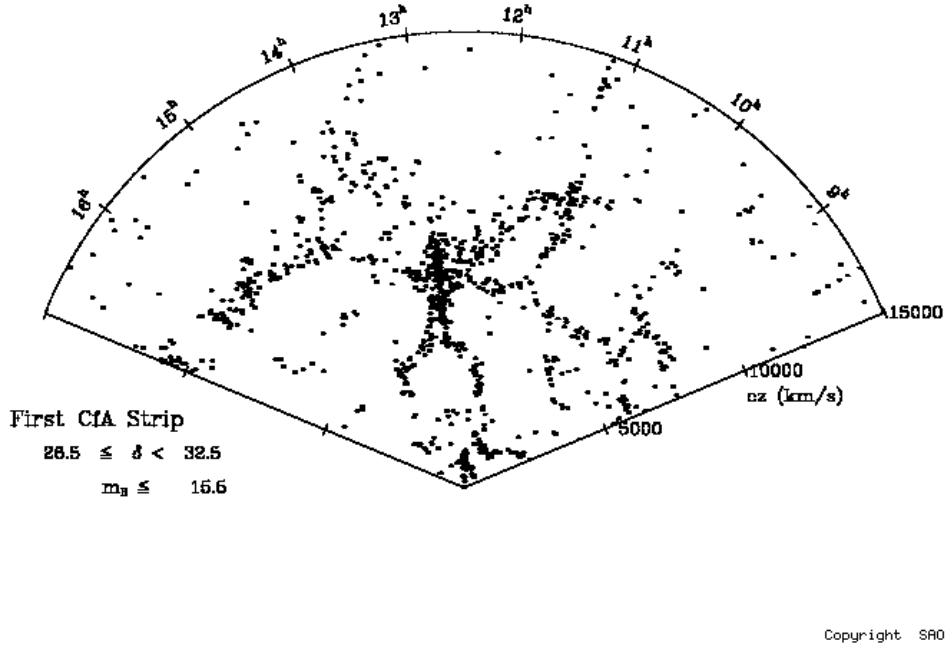


Figure 1.3: Galaxies are not uniformly distributed but form a bubble-like network as seen here in the CfA Redshift Survey ([de Lapparent et al., 1986]) where filaments of a few tens of megaparsecs join big clusters like the Coma cluster in the middle. The Coma Cluster looks elongated along the line-of-sight (creating the so-called "stickman" picture) because of redshift space distortions.

1.3 The growth of inhomogeneities

The hypothesis of cosmic homogeneity and isotropy if true can only be valid on extremely large scales. Indeed, large galaxy surveys have revealed the existence of superstructures (clusters of galaxies, superclusters and filaments and walls connecting them) on scales up to a few tens, as first seen by de Lapparent, Geller and Huchra in the CfA Redshift Survey ([de Lapparent et al., 1986]) – see figure 1.3 –, or even hundreds of megaparsecs (1 Mpc $\approx 3.1 \times 10^{22}$ m) such as the Great Wall discovered a few years later in the same catalogue ([Geller and Huchra, 1989]). The existence of large empty regions with size about 20 Mpc/h had already been suggested in the mid-seventies by [Gregory and Thompson, 1978, Jõeveer et al., 1978] who described them as large voids or holes with very low galaxy density between superclusters. Those ideas were reviewed in 1982 by Zel'dovich, Einasto and Shandarin under the title "Giant Voids in the Universe"([Zeldovich et al., 1982]).

More recent surveys like the Sloan Digital Sky Survey ([Abazajian et al., 2003]) or the 2MASS redshift survey ([Huchra et al., 2012]) have drastically improved our knowledge of the galaxy distribution showing with no doubt that galaxies form a complex web-like network on large scales made of voids, walls and filaments that interconnect with clusters of galaxies (see figure 1.4). This pattern is known as the *cosmic web* ([Klypin and Shandarin, 1993, Bond et al., 1996]).

In the standard paradigm of structure formation, the large-scale structures are believed to originate from initial quantum fluctuations that led to inhomogeneities in the CMB and in the matter density. Those initial density perturbations then grow under the laws of gravity and hierarchically form galaxies, clusters and super-clusters of galaxies. The next chapter will

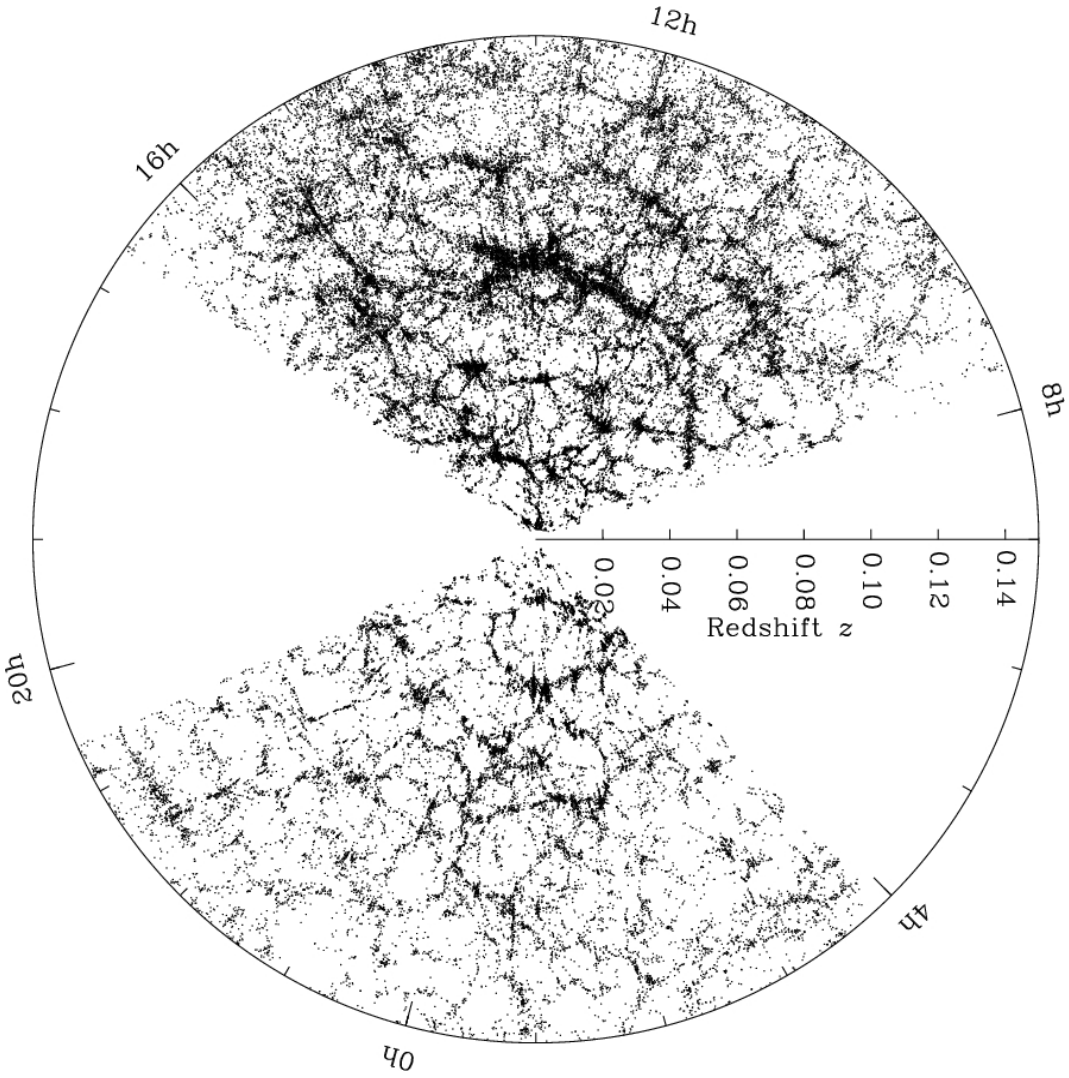


Figure 1.4: Catalogue of galaxies from the Sloan Digital Sky Survey. The cosmic web made of large voids, filaments and dense nodes is clearly visible.

be devoted to the theoretical understanding of the physical processes at play in forming the superstructures observed in the Universe.

2 Statistics of the large-scale structure

In this chapter, we review the cornerstones of the study of the large-scale structure of the Universe. More details are given in various textbooks including [Peebles, 1980, Padmanabhan, 1993, Peacock, 1999] and in the extensive review of [Bernardeau et al., 2002].

The large-scale structure of the Universe can be studied in a statistical sense as the result of the growth of primordial fluctuations under the laws of gravity. Those initial perturbations are characterized by their statistics – usually their power spectrum as they are observed to be almost Gaussian in the Cosmic Microwave Background (see figure 2.1). The temperature angular power spectrum of the CMB is shown on figure 1.2. Departure from Gaussianity are indeed shown [Planck Collaboration et al., 2014e] to be compatible with zero. Those fluctuations are then described by a random variable evolving under the gravity equations. We will first describe how gravitational instability leads to the formation of the large-scale structure of the Universe in section 2.1. Then section 2.2 will be devoted to the statistical description of the large-scale matter distribution. Finally, in section 2.3, we will describe the birth and growth of the so-called cosmic web and various ways of extracting this pattern from simulations and real data.

2.1 The gravitational instability

In the current cosmological paradigm, the mechanism that drives the growth of cosmic structures is gravitational instability : the densest regions attract matter from the rarest, meaning that peaks are becoming more peaky and voids are getting more empty. The starting point of the process is the primordial density field that is quasi-homogeneous on top of which tiny fluctuations grow until the formation of the cosmic structure that are seen in galaxy surveys.

The description of the growth of the primordial fluctuations can be made through Einstein's equations for an homogeneous universe perturbed by small density fluctuations δ. Will the appearance of a local over-density tend to disappear or to grow? In a medium of constant density, a local matter excess tends, thanks to its own gravitation, to attract the surrounding matter and therefore to grow. Following Jeans, this growth is exponential in a static space-time. However, in an expanding Universe, the expansion tends to dilute the fluctuations leading to a slower growth that is typically polynomial as we will show in the following.

Einstein's equations are highly non-linear but as far as sub-horizon modes are concerned, the problem can be reduced to its Newtonian approximation leading to the Vlasov-Poisson system described in section 2.1.1 Perturbation theory can then be applied for density contrasts much smaller than unity. The linear solution is well-known and described in section 2.1.2.2. It corresponds to a simple amplification of the fluctuations. Next-order contributions are increasingly difficult to predict as shown in section 2.1.2.3 and valid only in the weakly non-linear regime. For large densities ($\delta > 1$) where filaments and clusters form, the strongly non-linear regime can

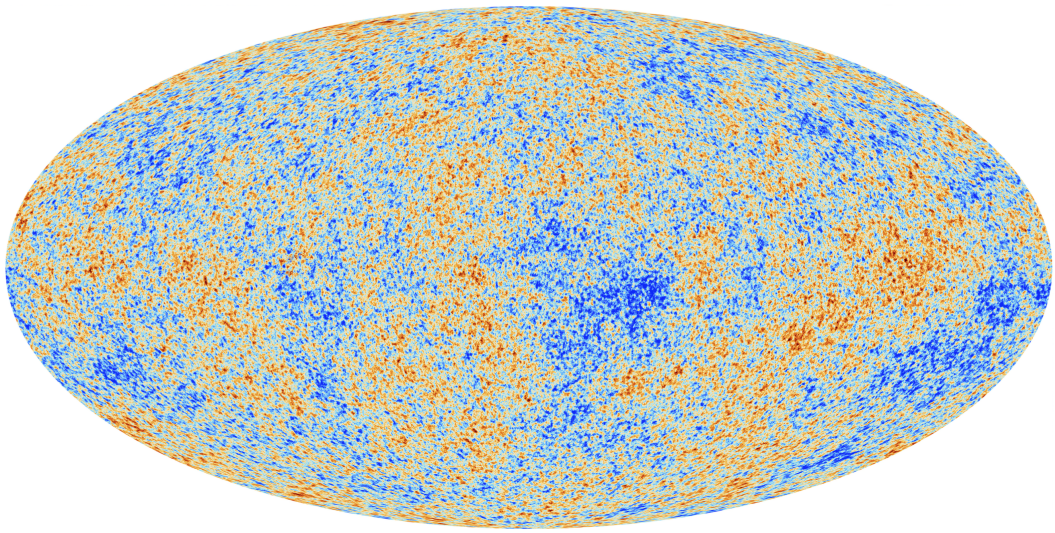


Figure 2.1: CMB temperature map measured by the Planck collaboration [Planck Collaboration et al., 2014a]. This field is shown to be almost Gaussian.

only be captured by numerical simulations (see section 2.1.4). An alternative is to study highly symmetric configurations that have analytical solutions (see section 2.1.3 for the spherical case)

2.1.1 The Vlasov-Poisson system

Let us study scales smaller than the horizon and much larger than the inter-particular distance. In this regime, the Newtonian approximation can be used to describe the evolution of a set of non-relativistic particles of mass m interacting via the gravitation force. In an expanding universe, we usually define the comoving coordinates $\mathbf{x}$ such that the physical distances are $\mathbf{r} = a\,\mathbf{x}$, the peculiar velocity $\mathbf{v}_p$ is the difference between the total velocity and the Hubble flow and the impulsion is $\mathbf{p} = m\,a\,\mathbf{v}_p$.

In phase space, the evolution of the phase space density $f(\mathbf{x}, \mathbf{p})$ is given by the Vlasov-Poisson system

$$\begin{cases} \dfrac{\mathrm{d}f}{\mathrm{d}t} = \dfrac{\partial}{\partial t} f(\mathbf{x},\mathbf{p}) + \dfrac{\mathbf{p}}{m\,a^2} \cdot \nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{p}) - m \nabla_{\mathbf{x}} \Phi(\mathbf{x}) \cdot \nabla_{\mathbf{p}} f(\mathbf{x},\mathbf{p}) = 0 \\ \Delta\Phi(\mathbf{x}) = \dfrac{4\pi G m}{a} \left(\displaystyle\int \mathrm{d}^3\mathbf{p}\, f(\mathbf{x},\mathbf{p}) - \frac{1}{V} \int \mathrm{d}^3\mathbf{x}\, \mathrm{d}^3\mathbf{p}\, f(\mathbf{x},\mathbf{p}) \right). \end{cases} , \tag{2.1}$$

where $\Delta = \nabla^2$ is the Laplacian and V is the comoving volume. By taking successive moments of the Vlasov equation with regards to the momentum, a full hierarchy of coupled equations is obtained. To resolve those equations, one needs to truncate the hierarchy at some order and add an additional hypothesis to close the system of equations. For instance, let us define the density field

$$\rho(\mathbf{x}) = \frac{m}{a^3} \int \mathrm{d}^3\mathbf{p}\, f(\mathbf{x},\mathbf{p}) \,, \tag{2.2}$$

the density contrast

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_0}{\rho_0} \,, \tag{2.3}$$

where ρ_0 is the mean density, the velocity field

$$\mathbf{u}(\mathbf{x}) = \frac{\int \mathrm{d}^3\mathbf{p}\,\mathbf{p}\,f(\mathbf{x},\mathbf{p})}{m\,a\int \mathrm{d}^3\mathbf{p}\,f(\mathbf{x},\mathbf{p})}\,, \tag{2.4}$$

and the velocity dispersion

$$\sigma_{ij}(\mathbf{x}) = \frac{\int \mathrm{d}^3\mathbf{p}\,\mathbf{p}_i\mathbf{p}_j f(\mathbf{x},\mathbf{p})}{m^2a^2\int \mathrm{d}^3\mathbf{p}\,f(\mathbf{x},\mathbf{p})} - \mathbf{u}_i\mathbf{u}_j\,. \tag{2.5}$$

Then, in addition to the Poisson equation, the first two moments of the Vlasov equation give the continuity and Euler equations

$$\begin{cases} \Delta\Phi(\mathbf{x}) = 4\pi G\rho_0 a^2\,\delta(\mathbf{x})\,, \\ \dfrac{\partial\delta}{\partial t} + \dfrac{1}{a}\nabla_{\mathbf{x}}\cdot[(1+\delta(\mathbf{x}))\mathbf{u}(\mathbf{x})] = 0\,, \\ \dfrac{\partial\mathbf{u}_i}{\partial t} + \dfrac{\dot{a}}{a}\mathbf{u}_i + \dfrac{1}{a}\mathbf{u}_j\mathbf{u}_{i,j} = -\dfrac{1}{a}\nabla_i\Phi - \dfrac{1}{\rho\,a}(\rho\sigma_{ij})_{,j}. \end{cases} \tag{2.6}$$

This system is closed if one assumes for instance that the fluid is thermalised $\sigma_{ij} = \delta_{ij}\frac{P}{\rho}$ and the pressure P is a simple function of the density through an equation of state.

2.1.2 Eulerian perturbation theories

2.1.2.1 Evolution of a single-flow fluid

Let us consider the gravitational evolution of a fluid in the single-flow regime. This approximation allows one to study the first stages of gravitational instability but breaks down as soon as shell-crossing occurs i.e when non-linear structures like filaments and halos form. In this case, the dynamics of the fluid is described by the following closed system of three coupled equations (Poisson equation, mass conservation and Euler equation)

$$\begin{cases} \Delta\Phi(\mathbf{x}) = 4\pi G\rho_0 a^2\,\delta(\mathbf{x})\,, \\ \dfrac{\partial\delta}{\partial t} + \dfrac{1}{a}\nabla_{\mathbf{x}}\cdot[(1+\delta(\mathbf{x}))\mathbf{u}(\mathbf{x})] = 0\,, \\ \dfrac{\partial\mathbf{u}_i}{\partial t} + \dfrac{\dot{a}}{a}\mathbf{u}_i + \dfrac{1}{a}\mathbf{u}_j\mathbf{u}_{i,j} = -\dfrac{1}{a}\nabla_i\Phi. \end{cases} \tag{2.7}$$

Those equations are highly non-linear but can be solved perturbatively as it will be shown in the two following sections.

2.1.2.2 Linear theory

At linear order, the hydrodynamical equations describe the evolution of the density contrast δ and the dimensionless divergence of the peculiar velocity $\theta = \frac{1}{aH}\nabla \cdot \mathbf{u}$[1] and read

$$\begin{cases} \dot{\delta} + H\theta = 0, \\ \dot{\theta} + 2H\theta + \dfrac{\dot{H}}{H}\theta = -\dfrac{4\pi G\rho_0}{H}\delta. \end{cases} \tag{2.8}$$

It is then straightforward to obtain the equation of evolution for the density contrast

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}H^2\Omega\delta = 0\,, \tag{2.9}$$

that has typically one growing and one decaying mode so that

$$\delta(\mathbf{x}, t) = D_+(t)\delta_+(\mathbf{x}, 0) + D_-(t)\delta_-(\mathbf{x}, 0)\,. \tag{2.10}$$

The solution for $\theta(\mathbf{x}, t)$ then reads

$$\theta(\mathbf{x}, t) = \frac{\partial \log D_+}{\partial \log a} D_+(t)\delta_+(\mathbf{x}, 0) + \frac{\partial \log D_-}{\partial \log a} D_-(t)\delta_-(\mathbf{x}, 0)\,. \tag{2.11}$$

For an Einstein-de Sitter universe, the solution is $D_+(t) \propto t^{3/2}$ and $D_-(t) \propto 1/t$ so that $\theta(\mathbf{x}, t) = \delta_+(\mathbf{x}, t) - 3/2\delta_-(\mathbf{x}, t)$.

2.1.2.3 PT kernels

Cosmological perturbation theory allows to go beyond linear theory and to study the effect of mode couplings on the gravitational dynamics. Starting from equations 2.7, we again neglect the rotational part of the velocity field and take the divergence of the Euler equation. Going to Fourier space [2], it yields

$$\begin{aligned} a\frac{\partial \delta(\mathbf{k}, a)}{\partial a} + \theta(\mathbf{k}, a) &= \frac{-1}{(2\pi)^{\frac{3}{2}}} \int \mathrm{d}^3\mathbf{k_1}\, \mathrm{d}^3\mathbf{k_2}\, \delta_D(\mathbf{k} - \mathbf{k_1} - \mathbf{k_2})\, \alpha(\mathbf{k_1}, \mathbf{k_2})\, \theta(\mathbf{k_1}, a)\, \delta(\mathbf{k_2}, a), \\ a\frac{\partial \theta(\mathbf{k}, a)}{\partial a} + \frac{1}{2}\theta(\mathbf{k}, a) + \frac{3}{2}\delta(\mathbf{k}, a) &= \frac{-1}{(2\pi)^{\frac{3}{2}}} \int \mathrm{d}^3\mathbf{k_1}\, \mathrm{d}^3\mathbf{k_2}\, \delta_D(\mathbf{k} - \mathbf{k_1} - \mathbf{k_2})\, \beta(\mathbf{k_1}, \mathbf{k_2})\, \theta(\mathbf{k_1}, a)\, \delta(\mathbf{k_2}, a), \end{aligned} \tag{2.12}$$

where α and β are defined as follows

$$\alpha(\mathbf{k_1}, \mathbf{k_2}) = 1 + \frac{\mathbf{k_1} \cdot \mathbf{k_2}}{k_1^2}$$

and

$$\beta(\mathbf{k_1}, \mathbf{k_2}) = \frac{\mathbf{k_1} \cdot \mathbf{k_2}}{2k_1^2} + \frac{\mathbf{k_1} \cdot \mathbf{k_2}}{2k_2^2} + \frac{(\mathbf{k_1} \cdot \mathbf{k_2})^2}{k_1^2 k_2^2}\,.$$

[1] It can be shown that in the no-shell crossing regime, the equation of evolution of the vorticity field does not have any source term, meaning that no vorticity can be created. On top of that, the vorticity is a decaying mode so that any initial vorticity will be diluted by the expansion.

[2] this is judicious as linearity in configuration space translates into mode independency in Fourier space

Note that Fourier transforms are defined so that for instance $\delta(\mathbf{k}) = \int \mathrm{d}^3\mathbf{x}\,\delta(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}/(2\pi)^{3/2}$. Assuming that the fields can be decomposed at all orders into initial fields

$$\delta(\mathbf{x},t) = \sum_n \delta^{(n)}(\mathbf{x},t),$$
$$\theta(\mathbf{x},t) = \sum_n \theta^{(n)}(\mathbf{x},t),$$

one can obtain a cosmological perturbation theory described at order n by

$$\delta^{(n)}(x) = \int \frac{\mathrm{d}^3\mathbf{k_1}}{(2\pi)^{3/2}}\delta(\mathbf{k_1})\ldots\frac{\mathrm{d}^3\mathbf{k_n}}{(2\pi)^{3/2}}\delta(\mathbf{k_n})\,a^n F_n(\mathbf{k_1},\ldots,\mathbf{k_n}),$$
$$\theta^{(n)}(x) = \int \frac{\mathrm{d}^3\mathbf{k_1}}{(2\pi)^{3/2}}\delta(\mathbf{k_1})\ldots\frac{\mathrm{d}^3\mathbf{k_n}}{(2\pi)^{3/2}}\delta(\mathbf{k_n})\,a^n G_n(\mathbf{k_1},\ldots,\mathbf{k_n}),$$

where F_n and G_n kernels can be computed recursively

$$F_1 = -G_1 = 1$$

and for $n \geq 2$:

$$F_n(\mathbf{k_1},\ldots,\mathbf{k_n}) = \sum_{i=1}^{n-1}\frac{G_i(\mathbf{k_1},\ldots,\mathbf{k_i})}{(2n+3)(n-1)}\Big[- (2n+1)\,\alpha(\mathbf{p_1},\mathbf{p_2})\,F_{n-i}(\mathbf{k_{i+1}},\ldots,\mathbf{k_n}) + 2\,\beta(\mathbf{p_1},\mathbf{p_2})\,G_{n-i}(\mathbf{k_{i+1}},\ldots,\mathbf{k_n})\Big],$$

$$G_n(\mathbf{k_1},\ldots,\mathbf{k_n}) = \sum_{i=1}^{n-1}\frac{G_i(\mathbf{k_1},\ldots,\mathbf{k_i})}{(2n+3)(n-1)}\Big[+ 3\,\alpha(\mathbf{p_1},\mathbf{p_2})\,F_{n-i}(\mathbf{k_{i+1}},\ldots,\mathbf{k_n}) - 2n\,\beta(\mathbf{p_1},\mathbf{p_2})\,G_{n-i}(\mathbf{k_{i+1}},\ldots,\mathbf{k_n})\Big],$$

with $\mathbf{p_1} = \mathbf{k_1} + ... + \mathbf{k_i}$, $\mathbf{p_2} = \mathbf{k_{i+1}} + ... + \mathbf{k_n}$. While all order of this series can be computed in principle, in practice each order is increasingly complex to compute so that it allows one at most to go to second or third order in the computation of low-order moments (mainly power spectra and bispectra). Note that all results obtained in this section were for an Einstein-de Sitter Universe for the sake of simplicity but it can be extended to other backgrounds (however the dependence on cosmology is typically very small).

2.1.2.4 Diagrammatic representation

The use of diagrams allows one to more easily handle perturbation theory orders. It is very similar to field theory with initial conditions, propagators and vertices. Indeed, it appears that dynamics equations can be rewritten in a simple matrix form using the vector $\Psi(\mathbf{k},\eta) = (\delta(\mathbf{k},\eta), -\theta(\mathbf{k},\eta))$ where η is a time variable such that $\eta = \ln a(\tau)$. Equations 2.12 yield

$$\partial_\eta\Psi_a(\mathbf{k},\eta) + \Omega_{ab}\Psi_b(\mathbf{k},\eta) = \gamma^{(s)}_{abc}(\mathbf{k},\mathbf{k_1},\mathbf{k_2})\Psi_b(\mathbf{k_1},\eta)\Psi_c(\mathbf{k_2},\eta) \tag{2.13}$$

with the usual conventions for repeated latin indices and integration on repeated Fourier variables. In equation 2.13, $\gamma^{(s)}_{abc}$ is the symmetrised vertex defined as

$$\gamma^{(s)}_{121}(\mathbf{k},\mathbf{k_1},\mathbf{k_2}) = \delta_D(\mathbf{k}-\mathbf{k_1}-\mathbf{k_2})\,\alpha(\mathbf{k_1},\mathbf{k_2})/2,$$
$$\gamma^{(s)}_{112}(\mathbf{k},\mathbf{k_1},\mathbf{k_2}) = \delta_D(\mathbf{k}-\mathbf{k_1}-\mathbf{k_2})\,\alpha(\mathbf{k_2},\mathbf{k_1})/2,$$
$$\gamma^{(s)}_{222}(\mathbf{k},\mathbf{k_1},\mathbf{k_2}) = \delta_D(\mathbf{k}-\mathbf{k_1}-\mathbf{k_2})\,\beta(\mathbf{k_1},\mathbf{k_2}),$$

and zero everywhere else and

$$\Omega = \begin{pmatrix} 0 & -1 \\ -3/2 & 1/2 \end{pmatrix}.$$

A solution of equation 2.13 can be found using a Laplace transform on η

$$\Psi_a(\mathbf{k},\eta) = g_{ab}(\eta)\phi_b(\mathbf{k}) + \int_0^\eta \mathrm{d}\eta' g_{ab}(\eta-\eta')\,\gamma^{(s)}_{bcd}(\mathbf{k},\mathbf{k_1},\mathbf{k_2})\Psi_c(\mathbf{k_1},\eta')\Psi_d(\mathbf{k_2},\eta'),$$

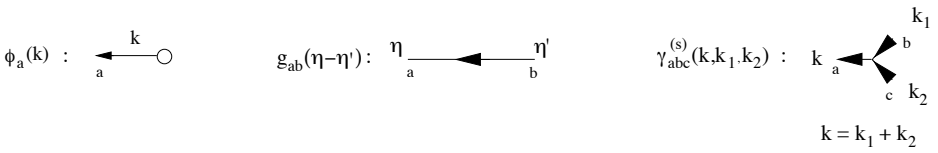
where $\phi_b(\mathbf{k}) = \Psi_b(\mathbf{k},\eta=0)$ is the initial condition for a scale factor $a(\tau)=1$ and g_{ab} is the linear propagator thus describing the linear evolution of density and velocity divergence fields. It is null by causality for $\eta<0$ and such that for $\eta>0$

$$g_{ab}(\eta) = \frac{e^\eta}{5}\begin{pmatrix} 3 & 2 \\ 3 & 2 \end{pmatrix} - \frac{e^{-3\eta/2}}{5}\begin{pmatrix} -2 & 2 \\ 3 & -3 \end{pmatrix}. \tag{2.14}$$

Equation 2.14 can be easily interpreted in terms of Feynman diagrams, the first term corresponding to the linear evolution of the initial conditions and the second term encoding all the non-linearities from mode couplings. Eventually, the theory is based on three ingredients to form diagrams:

- initial fields $\phi_a(\mathbf{k})$,
- a linear propagator $g_{ab}(\eta-\eta')$
- and three-leg vertices $\gamma^{(s)}_{abc}(\mathbf{k},\mathbf{k_1},\mathbf{k_2})$.

Those three bricks will be represented as follows



The non-linear terms carry all the interactions between pairs of wavevectors $\mathbf{k}_1$ and $\mathbf{k}_2$ at all times η' between 0 and η such that the sum $\mathbf{k}_1+\mathbf{k}_2$ is $\mathbf{k}$ (the so-called mode couplings). This interaction is characterized by the vertex $\gamma^{(s)}_{abc}(\mathbf{k},\mathbf{k_1},\mathbf{k_2})$ and the resulting wave evolves *linearly* following the propagator $g_{ab}(\eta-\eta')$. Therefore, to get the order-n diagrams, one has to draw all possible trees with $n-1$ vertices and n initial conditions. Starting from a leaf labelled by the time η, one can go back in time until a vertex that gives rise to two different branches which can be followed until either another vertex that itself creates two other branches, or an initial field at time 0. To interpret the diagram, one has to label all branches by a wavevector $\mathbf{k}_i$ (so that there is momentum conservation at each vertex), label all vertices by a time s_j, count each vertex, each branch (and attribute a free propagator for k_i between s_j and s_k) and each initial field and finally sum over all times and intermediate momenta. For symmetry reasons,

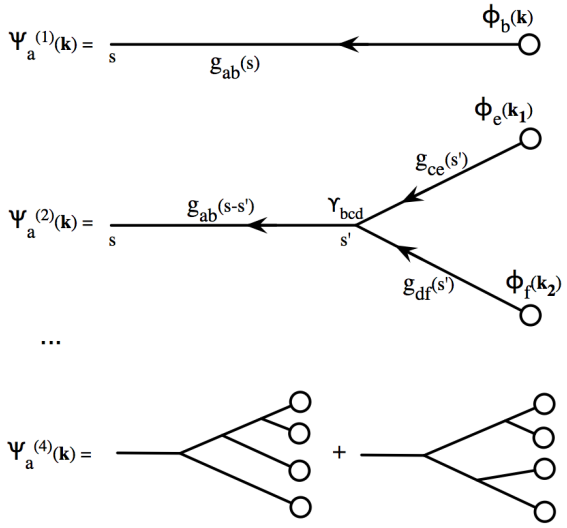


Figure 2.2: Diagrammatic representation of the field Ψ to fourth order.

each non-symmetric configuration must be counted twice. One example of such a diagram is displayed in figure 2.2.

For instance, the density field at second order $\Psi_1^{(2)}(\mathbf{k}, s)$ can be computed starting from the left-hand side of the diagram (i.e. from the time s). $g_{1b}(s - s')$ for the propagation between s and s', γ_{bcd} for the vertex at s', $g_{ce}(s' - 0) \times \phi_e(\mathbf{k_1})$ for the upper branch and $g_{df}(s' - 0) \times \phi_f(\mathbf{k_2})$ for the other one, have to be integrated for all possible s' between 0 and s and intermediate wavevectors $\mathbf{k}_1$ et $\mathbf{k}_2$ such that $\mathbf{k_1} + \mathbf{k_2} = \mathbf{k}$. In the end, it gives the following expression for the density field at second order

$$\Psi_1^{(2)}(\mathbf{k}, s) = \int \mathrm{d}s' \, \mathrm{d}^3\mathbf{k_1} \, \mathrm{d}^3\mathbf{k_2} \, g_{1b}(s - s')\gamma_{bcd} \, g_{ce}(s') \, \phi_e(\mathbf{k_1}) \, g_{df}(s') \, \phi_f(\mathbf{k_2}) \, \delta_D(\mathbf{k} - \mathbf{k_1} - \mathbf{k_2}) \, .$$

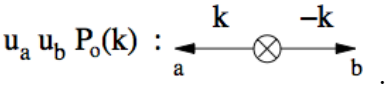
This formalism allows one to compute the successive moments of the density and velocity divergence fields. Assuming that the initial fields are Gaussian, one can write

$$\phi_a(\mathbf{k}) = u_a \delta_0(\mathbf{k}),$$

with $u_1 = u_2 = 1$ for a pure growing mode in the initial conditions. Those initial conditions are characterized by the initial power spectrum

$$\langle \phi_a(\mathbf{k}) \, \phi_b(\mathbf{k}') \rangle = \delta_D(\mathbf{k} + \mathbf{k}') \, u_a \, u_b \, P_0(\mathbf{k}),$$

which can be represented as



The idea is now to investigate how this initial statistics evolves under gravity and how non-Gaussianities arise and skew the probability density function. To do so, let us first consider the power spectrum at time η

$$\langle \Psi_a(\mathbf{k}, \eta) \Psi_b(\mathbf{k}', \eta) \rangle = \delta_D(\mathbf{k} + \mathbf{k}') P_{ab}(\mathbf{k}, \eta).$$

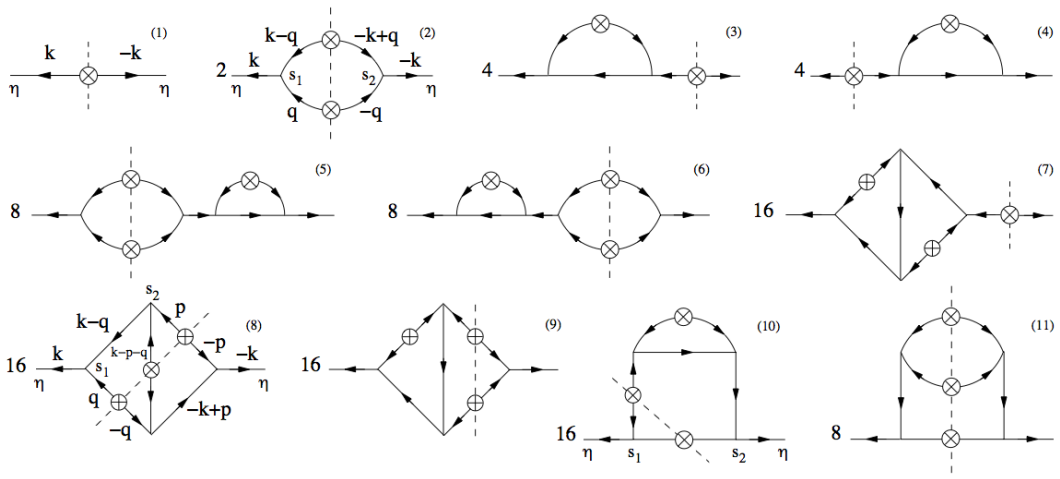


Figure 2.3: Part of the diagrams of $P_{ab}(\mathbf{k},\eta)$ until two-loop order. Dashed lines show the location where the initial fields of $\Psi_a^{(m)}$ and $\Psi_a^{(2l+2-m)}$have been glued together.

This power spectrum can be expressed as the sum

$$P_{ab}(\mathbf{k},\eta) = \sum_{l=0}^{\infty} P_{ab}^{(l)}(\mathbf{k},\eta)$$

of the l-loop power spectra

$$\delta_D(\mathbf{k}+\mathbf{k}')P_{ab}^{(l)}(\mathbf{k},\eta) = \sum_{m=1}^{2l+1} \langle \Psi_a^{(m)}(\mathbf{k},\eta)\Psi_b^{(2l+2-m)}(\mathbf{k}',\eta)\rangle .$$

In order to draw each diagram of the p-loop power spectrum, one has to choose m between 1 and $2p+1$, put one tree diagram of $\Psi_a^{(m)}$ and one tree diagram of $\Psi_a^{(2p+2-m)}$ face to face and glue their initial fields together into an initial power spectrum $\delta_D(\mathbf{k_i}+\mathbf{k_j})u_c u_d P_0(\mathbf{k_i})$, the crucial point being to properly take into account all combinatory factors. Figure 2.3 illustrates this procedure. For instance, the contribution from the second diagram to the power spectrum reads

$$P_{ab}^{\mathrm{diag}(2)} = \int \mathrm{d}^3\mathbf{q}\,\mathrm{d}s_1\,\mathrm{d}s_2\, g_{ac}(\eta-s_1)\,\gamma_{cde}(\mathbf{k},\mathbf{q},\mathbf{k}-\mathbf{q})\,g_{df}(s_1)\,u_f\,g_{eg}(s_1)\,u_g\,P_0(q)\,P_0(|\mathbf{k}-\mathbf{q}|)$$
$$\times\quad g_{bh}(\eta-s_2)\,\gamma_{hij}(-\mathbf{k},-\mathbf{q},\mathbf{q}-\mathbf{k})\,g_{ik}(s_2)\,u_k\,g_{jl}(s_2)\,u_l\,. \quad (2.15)$$

This standard theory of cosmological perturbations is based upon an expansion in powers of the density field. At tree order and in the large-scale limit (where perturbations are very small), this expansion is well-justified but as soon as non-linearities arise through p-loop corrections, this approximation breaks down. Indeed, perturbation theory is not based on a small coupling constant but the expansion depends in particular on scale and redshift. In particular, the theory suffers from a lack of convergence as higher order corrections tend to oscillate instead of being purely additive. However, pushing the predictions beyond linear order is crucial in cosmology in order to reach the precision required to describe the statistics of the large-scale structure (e.g baryon acoustic oscillations in the power spectrum). Several attempts have thus been made to try and cure the convergence problems. One of these propositions is to renormalize the theory in terms of partial resummations of the series that allow to have a constructive perturbation theory in the sense that each contribution is positive. This renormalized perturbation theory is described in [Crocce and Scoccimarro, 2006, Bernardeau et al., 2008, Bernardeau et al., 2010]

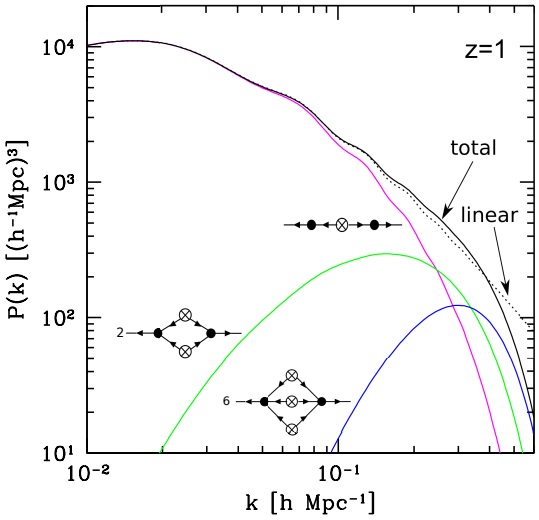


Figure 2.4: Constructive contributions to the density power spectrum at 0 (purple), 1 (green) and 2-loop (blue) in the RegPT approach and comparison with the linear prediction (dotted line). This figure is from [Taruya et al., 2012].

and can be regularised at low-k in the so-called RegPT approach ([Taruya et al., 2012]). An illustration of the density power spectrum predicted by RegPT is shown in figure 2.4. Other approaches include time renormalisation group approach ([Pietroni, 2008, Lesgourgues et al., 2009]), closure theory ([Taruya and Hiramatsu, 2008, Hiramatsu and Taruya, 2009]), effective field theory ([Carrasco et al., 2012]).

2.1.3 An exact non-linear solution : the spherical collapse

In the hierarchical model, objects of a few million solar masses (star clusters) form first and then gather by gravitational instability to form galaxies which next form clusters and superclusters. Their characteristics can be studied in the frame of the spherical model which we now develop. It allows one to understand how, for instance, a cluster of galaxies can break away from the cosmic expansion.

Let us assume the existence of a spherical region of size R and spatially constant density ρ. Birkhoff's theorem dictates the evolution of the radius of the overdensity:

$$\ddot{R} = -\frac{GM}{R^2} \quad \text{where } M = \frac{4}{3}\pi\left(\rho - \frac{\Lambda}{8\pi G}\right)R^3.$$

Equivalently, energy conservation can be written: $1/2\dot{R}^2 - GM/R = E$. If E is negative, the system is bound and the solution is

$$\frac{R}{R_m} = \frac{1}{2}(1-\cos\eta)\,,$$

$$\frac{t}{t_m} = \frac{1}{\pi}(\eta - \sin\eta)\,,$$

meaning that the radius R grows until $R_m = -GM/E$ when $t = t_m = \pi GM/\sqrt{8}(-E)^{3/2}$ and then shrinks until $R = 0$ for $t = 2t_m$. It is then possible to write the equation of evolution of R and t for $\eta \ll 1$ – meaning for the initial evolution of the overdensity – and find the first terms of the development of $R(t)$

$$R(t) = \frac{R_m}{4}\left(\frac{6\pi t}{t_m}\right)^{\frac{2}{3}}\left(1 - \frac{1}{20}\left(\frac{6\pi t}{t_m}\right)^{\frac{2}{3}} + ...\right).$$

Let us consider an overdensity of initial radius R_0 in an Einstein - de Sitter Universe of background density ρ_0, after a time t, $\rho = \rho_0(1+\delta)$ and the new radius of the overdensity is R, then as there is mass conservation

$$\frac{4}{3}\pi\rho_0 R_0^3 = \frac{4}{3}\pi\rho_0(1+\delta)R^3,$$

so that

$$R = R_0(1+\delta)^{-\frac{1}{3}},$$

which becomes, to first order,

$$R = R_0\left(1 - \frac{\delta}{3}\right).$$

Hence, the mean overdensity compared to an Einstein-de Sitter Universe is $\delta = 3/20\,(6\pi t/t_m)^{\frac{2}{3}}$. The non-linear collapse of the sphere ends at $R = 0$ and $t = 2t_m$ which corresponds to an extrapolated linear density of

$$\delta_{\rm collapse} = \delta(2t_m) \approx 1.686 = \delta_c\,.$$

This critical density will be useful in the Press-Schechter formalism as we often assume that the entire non-linear collapse occurs when the linear density equals this critical density. A spherical peak or void can be described in this spherical model as concentric mass shells which never cross during the collapse (actually it depends on the initial density profile and this description, in terms of spherical collapse, is valid until first shell-crossing). Each shell acts as an independent Friedmann universe and we can derive the dynamical equations. Note that the energy E can be derived as the sum of the initial ($t_i = 0$) kinetic K and potential U energies so that $E = H^2(t_i)R_i^2(1 - \Omega_i(1+\delta_i))/2$ where Ω_i is the initial ratio between the background density and the critical density. As, the mean density within one shell is given by $3M/4\pi R^3$, mass conservation within one shell together with the dynamical evolution of the background density yield the equation of evolution of the density contrast. A solution is known for $\Lambda = 0$ ([Peebles, 1980]). If $\delta = \rho/\rho_0 - 1$ is the density contrast, this solution can be easily written for an Einstein - de Sitter Universe $\Omega = 1$

- for an initial overdensity ($\delta_i > 0$),

$$\delta_i \frac{a(t)}{a(t_i)} = \frac{3}{5}\left(\frac{3}{4}(\eta - \sin\eta)\right)^{2/3}$$

$$1 + \delta = \frac{9}{2}\frac{(\eta - \sin\eta)^2}{(1 - \cos\eta)^3};$$

- for an initial underdensity ($\delta_i < 0$),

$$\delta_i \frac{a(t)}{a(t_i)} = -\frac{3}{5}\left(\frac{3}{4}(\sinh\eta - \eta)\right)^{2/3}$$

$$1+\delta = \frac{9}{2}\frac{(\sinh\eta-\eta)^2}{(\cosh\eta-1)^3}.$$

In particular, underdense regions are increasingly emptying and vice versa. For $\Omega \to 0$, the solution reads

$$1+\delta = \frac{1}{(1-2\frac{\epsilon}{3})^{3/2}},$$

where ϵ is the linear contrast $\epsilon = D_+(t)\delta_i$. Note that when the spherical collapse solution is expressed as a function of the linear contrast ϵ, the dependence on cosmology is very weak as shown in [Bernardeau et al., 2002].

The next step is to study the collapse of a homogeneous ellipsoidal overdensity to describe anisotropic objects like filaments or walls. Triaxial overdensities tend to evolve more quickly along their shortest axis. Hence, the shortest axis tend to collapse first while the others continue to grow forming flattened structures which are often called pancakes. Notice that this continuous description does not consider substructures. Nevertheless, it allows one to understand the physical processes which lead to the observed large-scale structure of the Universe. Obviously real objects will not collapse until a point of infinite density but will reach virial equilibrium after turn-around. This equilibrium takes place when the total energy is half the potential energy ($E = U(R) + K(R) = U(R_{\rm vir})/2$). As $E = U(R_m)$, the virial radius $R_{\rm vir}$ is half the radius at turn-around R_m, which corresponds to an overdensity at virialization about $\delta_{\rm vir} = 178$. This criterion is traditionally used to define collapsed objects.

Beyond the qualitative description of halo formation, the spherical collapse ([Gunn and Gott, 1972]) is essential to derive analytical results based on perturbation theory. This aspect will be developed in more details in section 5.1.2.

2.1.4 Numerical simulations

To study the whole non-linear gravitational evolution of the Universe, one is led to consider numerical simulations. Such simulations are based on two main ingredients: initial conditions (that represent the Universe as seen in the cosmic microwave background) and a physical theory to evolve the initial conditions across cosmic time. In the end, it allows one to build robust predictions that are eventually compared to observations of our Universe in order to study the physics governing the birth and growth of cosmic structures. Typically, two types of simulations are used to study galactic and extra-galactic physics: N-body codes, where matter density fluids are discretized in particles, and hydrodynamical codes.

2.1.4.1 N-body simulations

The first N-body simulations in astrophysics were made in the sixties. In the seventies, the first simulations of galaxy clusters were run with a few tens of particles ([Peebles, 1970, White, 1976]). The first cosmological simulations came quickly after ([Press and Schechter, 1974, Aarseth et al., 1979, Doroshkevich et al., 1980, Efstathiou and Eastwood, 1981a]) with a few thousands particles leading to the theoretical modeling of the cosmic web ([Centrella and Melott, 1983, Frenk et al., 1983]).

Starting from the Vlasov-Poisson system (equation 2.1), the equations describing the evolution of a finite set of particles with comoving coordinate $\mathbf{x}$ and peculiar velocity $\mathbf{v}$, and tracing the

matter distribution, simply read

$$\mathbf{v} = a\dot{\mathbf{x}}, \tag{2.16}$$

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + H\mathbf{v} = -\frac{\nabla\Phi}{a}\,. \tag{2.17}$$

The evolution of $a(t)$ is set by the background evolution i.e by the Friedmann equations once the cosmological parameters are fixed. The solution to this N-body problem then relies on a numerical scheme. Some of these schemes are described below. The generation of initial conditions for cosmological simulation is also a problem on its own. The basic idea is to generate a Gaussian perturbation field with a power spectrum given by equation 2.27 and compute the corresponding displacement field. Particles initially located on the grid are then assigned an initial velocity given by the Zel'dovich approximation (see also section 2.3.1). Note that improvements of this method are numerous. First, particles can be located on an amorphous fully relaxed distribution instead of a regular grid, the advantage being to avoid preferred directions ([White, 1996]). Then, one can go beyond the Zel'dovich approximation to assign initial velocities and use Lagrangian perturbation theory at second order ([Jenkins, 2010]). This allows one, in particular, to start the simulation at a later redshift.

The particle-particle (PP) method The simplest, but most time-consuming, method to solve the N-body problem is to compute, at each time step, the exact Newtonian forces that act between two particles of mass m_i and m_j separated by r_{ij}

$$F_{ij} = \frac{Gm_i m_j}{r_{ij}^2} \tag{2.18}$$

so that each particle has a new velocity and position given by

$$\mathbf{v}(t_i + \Delta t) = \mathbf{v}(t_i) + \frac{\sum_j F_{ij}\Delta t}{m_i}\,, \tag{2.19}$$

$$\mathbf{x}(t_i + \Delta t) = \mathbf{x}(t_i) + \mathbf{v}(t_i + \Delta t)\Delta t\,. \tag{2.20}$$

The main drawback of this approach is that it scales like the square of the number of particles N_p^2 making the resolution very difficult when the number of particles become large. Note that in practice, a softening parameter ϵ is introduced to cure the divergence of the gravitational potential

$$\phi = -\frac{Gm}{\sqrt{r^2 + \epsilon^2}}\,. \tag{2.21}$$

This can be interpreted as the finite size of each particle. In particular, it reduces the spurious effect of two-body relaxation. Note also that the accuracy of the PP method resides in the smallness of the timestep and in the integrator used. Different schemes have been proposed so far.

The particle-mesh (PM) method Another approach to compute the gravitational forces is to directly integrate the Poisson equation on a mesh and get the gravitational potential in Fourier space where Fast Fourier Techniques can be efficiently used. The execution time is drastically reduced but the accuracy of the method is also degraded in particular on small scales. In more detail, at each time step, the density on the grid is computed, then Poisson's equation is solved to find the gravitational potential, forces are derived from the potential on each grid point and interpolated to each particle position before the equations of motion are integrated.

Different interpolation schemes can be used to evaluate the density field on the mesh (and then to interpolate consistently the force at the particle locations): "NGP" where the mass of the particle is assigned to the nearest grid point, "cloud-in-cell" where the mass is shared between the 2^D nearest points depending on the inverse distance between them, etc. The solution of the Poisson equation is the most time-consuming step but Fast Fourier Techniques are very efficient so that the PM method eventually scales like $N_p + N_{\rm grid} \log N_{\rm grid}$ depending on the number of particles N_p and grid points $N_{\rm grid}$. Hybrid codes between PM and PP methods have been developed in order to keep enough resolution on small scales (PP) and be efficient on large scales (PM). They are called P^3M ([Efstathiou and Eastwood, 1981b]) or AP^3M ([Couchman, 1991]) when they are in addition adaptative i.e when the mesh is refined in large-density regions.

Tree codes Tree codes ([Appel, 1985, Barnes and Hut, 1986]) are now the most widely used methods to solve N-body problems in cosmology. They are based on a tree algorithm which describes the matter distribution as a tree. First, a large cubic cell is built that encompasses all particles. It is then recursively subdivided into eight sub-cells until each cell contain one and only one particle. The force is then computed by descending the tree and applying a hierarchical multipole expansion i.e if the angular opening of the cell is less than a fixed acceptation angle Θ, the cell is considered as a single particle else sub-cells are considered. Depending on the acceptation angle, this method can be as efficient as $\mathcal{O}(N_p \log N_p)$ ($\Theta \approx 0.7$ rad).

Hybrid methods called Tree-PM that mixes a long-range PM method with tree code on small scale are very efficient and allows one to run large simulations with very good resolution. Gadget ([Springel, 2005]) is one of them.

Recent simulations It is now possible to simulate the evolution of self-gravitating collisionless matter (e.g cold dark matter) in cosmological volumes (of several billions light year across) with a huge dynamical range. For instance, the Horizon 4π simulation ([Teyssier et al., 2009]) maps half the observable universe, with enough resolution to describe a Milky Way-like galaxy with more than 100 dark matter particles. Let me also mention the Millenium simulations ([Springel et al., 2005, Angulo et al., 2012]) run with Gadget and widely used in the community, the Bolshoi simulation ([Klypin et al., 2011]) run with the AMR code Art, the Horizon Run 3 ([Kim et al., 2011]) or the DEUS Full Universe simulations ([Alimi et al., 2012]) that use different cosmological models (including different models of dark energy like Ratra-Peebles quintessence or phantom fluid). Recently, the first codes trying to resolve the Vlasov-Poisson system directly in phase space have been developed ([Hahn and Angulo, 2015], [Colombi et al., 2015]). Those codes allow one to trace with much more accuracy phase space foldings during shell crossings but are computationally costly as they require one to follow the evolution of a 3D sheet in a 6D space. For cold dark matter, they give confidence in the standard N-body solvers that reproduce with surprising accuracy the results obtained with Vlasov-Poisson solvers. It seems that dark matter simulations have therefore converged today and are able to provide robust predictions for some dark matter halo statistics. This conclusion is far from being reached by hydrodynamical simulations as it will be discussed in section 2.1.4.2.

The Horizon-4π simulation In this manuscript, we will in particular make use of the Horizon 4π N-body simulation ([Teyssier et al., 2009]). This simulation which contains 4096^3 DM particles distributed in a 2 h^{-1}Gpc periodic box is characterized by the following ΛCDM cosmology: $\Omega_{\rm m} = 0.24$, $\Omega_\Lambda = 0.76$, $n = 0.958$, $H_0 = 73$ km$\cdot s^{-1}\cdot$Mpc^{-1} and $\sigma_8 = 0.77$ within one standard

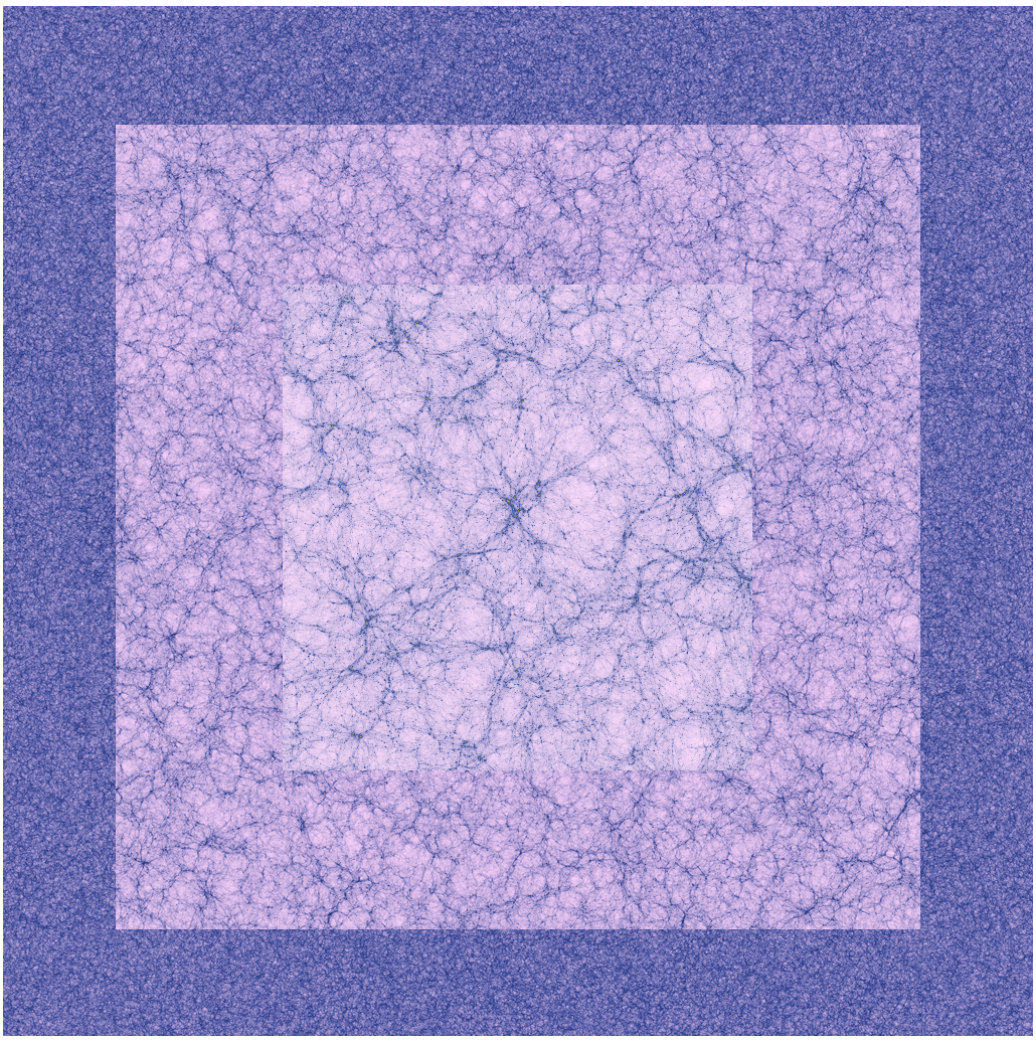


Figure 2.5: The Horizon 4π simulation. The outer region corresponds to a view of the universe on scales of $16h^{-1}$Gpc: it is generated by unfolding the simulation while cutting a slice obliquely through the cube in order to preserve the continuity of the field (thanks to the periodicity). The intermediate region corresponds to a slice of $2h^{-1}$Gpc, while the inner region is at the original resolution of the initial conditions. Credits: The Horizon project.

deviation of WMAP3 results ([Spergel et al., 2003]). These initial conditions were evolved non-linearly down to redshift zero using the adaptive mesh refinement code Ramses ([Teyssier, 2002]), on a 4096^3 grid. The motion of the particles was followed with a multi grid Particle-Mesh Poisson solver using a Cloud-In-Cell interpolation algorithm to assign these particles to the grid (the refinement strategy of 40 particles as a threshold for refinement allowed us to reach a constant physical resolution of 10 kpc, see the above references). The Friend-of-Friend (FoF) Algorithm ([Huchra and Geller, 1982]) was used over 18^3 overlapping subsets of the simulation with a linking length of 0.2 times the mean interparticular distance to define dark matter haloes. Only haloes with more than 40 particles are typically considered, which corresponds to a minimum halo mass of $3 \times 10^{11} M_\odot$ (the particle mass being $7.7 \times 10^9 M_\odot$). The mass dynamical range of this simulation spans about 5 decades. Overall, 43 million halos were detected at redshift zero.

2.1.4.2 Simulations of baryons and dark matter

On large scales, the effect of baryonic processes is very small and baryons simply follow the dark matter dynamics. However, on small-to-intermediate scales, baryonic physics cannot be neglected and should be taken into account. First ideas to study the baryonic counterpart of dark matter simulations were to rely on dark matter simulations and use recipes to populate dark matter halos with galaxies. These semi-analytical models are nonetheless not sufficient to account for the importance of non-linear gas dynamics and baryonic processes. Hydrodynamical simulations however are able to take into account ab initio these complex baryonic phenomena. They can be split into two types of codes: SPH and AMR that are described below.

Semi-analytical models Dark matter only simulations are much faster to run that hydrodynamical ones. However, in order to be compared to observations, they need to be post-processed and populated with galaxies of different sizes and luminosities. This can be done with semi-analytical models ([Kauffmann et al., 1999, Springel et al., 2001, Croton et al., 2006, Baugh, 2006]) or halo occupation distribution models ([Seljak, 2000, Ma and Fry, 2000, Peacock and Smith, 2000, Scoccimarro et al., 2001, Berlind and Weinberg, 2002]). If they are successful to reproduce some observations, there is a real need to model self-consistently baryons and dark matter and therefore to rely on hydrodynamical simulations.

Hydrodynamical simulations Baryons are usually described as a perfect fluid that follows the standard hydrodynamical equations namely Euler and continuity equations and the first law of thermodynamics that can be written in an expanding Universe

$$\begin{cases} \dfrac{\partial \mathbf{v}}{\partial t} + \dfrac{1}{a}(\mathbf{v}\cdot\nabla)\mathbf{v} + \dfrac{\dot{a}}{a}\mathbf{v} = -\dfrac{1}{a\rho}\nabla P - \dfrac{1}{a}\nabla\Phi\,, \\ \dfrac{\partial \rho}{\partial t} + 3\dfrac{\dot{a}}{a}\rho + \dfrac{1}{a}\nabla\cdot(\rho\mathbf{v}) = 0\,, \\ \dfrac{\partial(\rho U)}{\partial t} + \dfrac{1}{a}\mathbf{v}\cdot\nabla(\rho U) = -(\rho U + P)\left(\dfrac{1}{a}\nabla\cdot\mathbf{v} + 3\dfrac{\dot{a}}{a}\right)\,, \end{cases} \tag{2.22}$$

where P is the pressure and U the internal energy per unit mass. For a monoatomic gas those quantities are related through $P = 2/3\rho U$. The joint evolution of baryons and dark matter is often solved using the N-body techniques described in section 2.1.4.1 to model the gravitational part and specific hydrodynamical methods are designed for the baryonic part. Typically two

main approaches can be followed: smoothed particle hydrodynamics (Lagrangian view) and grid codes (Eulerian view).

In smooth particle hydrodynamical codes (SPH), the fluid is discretized into particles. If the major advantage of such a method is that it is Lagrangian, hence it refines naturally in high density environment; on the other hand shocks and low-density environment are more difficult to capture accurately. In order to mitigate the difficulty of describing discontinuities with smooth particles (self-gravity in cosmological setups induces high complexity like supersonic motions, shocks and discontinuities), they often need to introduce an artificial viscosity. The first implementation in cosmology was made by [Evrard, 1988]. Examples of such Lagrangian codes are e.g Gadget ([Springel, 2005]) and Gasoline ([Wadsley et al., 2004]).

The opposite point of view is to solve equations 2.22 on a (fixed or adaptative) grid. Quantities inside each cells (like density, temperature, pressure, etc) are updated at each time step once fluxes are computed for all interfaces. Those methods are very accurate in low and large-density environments and in describing shocks. Such Eulerian codes include Ramses ([Teyssier, 2002]), Enzo ([Bryan et al., 1995, O'Shea et al., 2004]), Flash ([Fryxell et al., 2000]), etc.

Hybrid codes like moving mesh are designed to gather the best of the Eulerian and Lagrangian method (high resolution together with precise shock description) but are difficult to implement in practice and their error budget is difficult to carry out. They have been shown to behave closely to AMR codes. A recent example of moving mesh code is Arepo ([Springel, 2010]).

Already at the level of the numerical codes to solve the hydrodynamics, discrepancies – sometimes quite large – between different approaches are found ([Scannapieco et al., 2012]). On top of that, galaxy formation simulations must also account for non-adiabatic processes such as cooling and heating and must deal with sub-grid physics and feedbacks as detailed below.

Baryonic processes In addition to gravity and gas dynamics, lots of baryonic processes have to be taken into account. Some of them have drastic impacts on the physical results and are central in the study of cosmic structures and galaxy formation. A non-exhaustive list is given here.

- radiative gas cooling and photoheating from a UV background : In a plasma of hydrogen and helium, collisions induce excitation, ionisation, recombination and finally free-free emissions. When optically thin gas and ionisation equilibrium are assumed, the resulting cooling function $\Lambda(\rho, U)$ can be computed or extracted from a table and added to the first law of thermodynamics.
- star formation : In cold and dense regions, stars form. This sub-grid phenomenon is implemented through a conversion of a fraction of the gas into collision-less star particles that have the same dynamical behaviour as dark matter particles. The most widely used recipe is the Schmidt law for which the star formation rate is proportional $\rho^{3/2}$. The efficiency of star formation is a parameter of the simulation together with a minimum star particle mass.
- stellar feedback : it includes UV photons that may couple to the gas component via radiation pressure, mass loss by stellar winds and supernovae explosions for which energy can be deposited thermally and kinetically. As those events occur at the parsec scale that is not resolved by large-scale structure simulations, a sub-grid recipe is required. Various attempts have been deployed, the first ones led to the so-called cooling catastrophe where too many stars form. Different models have then been proposed that could prevent from

this catastrophe but they still remain ad hoc prescriptions that need a better understanding of the processes driving galaxy formation. This feedback is important as it is energetic enough to impede galaxy formation at low mass.

- Active Galactic Nuclei (AGN) feedback : supermassive black holes at the centers of galaxies can have a drastic impact on their large-scale environment through kinematically energetic outflows and relativistic jets. It is the current preferred scenario that could explain the properties of massive galaxies.
- chemical enrichment
- thermal conduction
- radiation pressure from young massive stars
- magnetic fields
- etc.

Eventually, galaxy formation simulations have lots of sources of uncertainties coming from the hydrodynamical code used, the recipes that model sub-grid physics and the number of parameters.

Recent simulations State-of-the-art cosmological hydrodynamical simulations include the Mare Nostrum simulation ([Ocvirk et al., 2008]), the Horizon-AGN simulation (see figure 2.6) run with the adaptative mesh refinement code Ramses (see below), Illustris ([Vogelsberger et al., 2014]) run with the Arepo code, Massive Black II ([Khandai et al., 2015]) or EAGLE ([Schaye et al., 2015]) run with upgraded versions of Gadget3.

The Horizon-AGN simulation The hydrodynamical simulation used in this manuscript is the Horizon-AGN simulation ([Dubois et al., 2014]). A standard ΛCDM cosmology compatible with the WMAP-7 cosmology ([Komatsu et al., 2011]) is adopted, with total matter density $\Omega_{\rm m} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, amplitude of the matter power spectrum $\sigma_8 = 0.81$, baryon density $\Omega_{\rm b} = 0.045$, Hubble constant $H_0 = 70.4\,{\rm km\,s^{-1}\,Mpc^{-1}}$, and $n_s = 0.967$. The Horizon-AGN simulation has been run with 1024^3 dark matter (DM) particles in a $L_{\rm box} = 100\,h^{-1}\,{\rm Mpc}$ box, so as to obtain a DM mass resolution of $M_{\rm DM,res} = 8 \times 10^7\,{\rm M_\odot}$. The Adaptive Mesh Refinement code ramses ([Teyssier, 2002]) has been used to run the simulation with an initial mesh refinement of up to $\Delta x = 1\,{\rm kpc}$ (7 levels of refinement). The refinement scheme follows a quasi-Lagrangian criterion: if the number of DM particles in a cell is more than 8, or if the total baryonic mass in a cell is 8 times the initial DM mass resolution, a new refinement level is triggered.

A [Sutherland and Dopita, 1993] model is used to allow gas cooling by means of H and He cooling down to $10^4\,{\rm K}$ with a contribution from metals. Following [Haardt and Madau, 1996], heating from a uniform UV background takes place after redshift $z_{\rm reion} = 10$. Metallicity is modelled as a passive variable for the gas that varies according to the injection of gas ejecta during supernovae explosions and stellar winds. A Schmidt law is used to model star formation: $\dot\rho_* = \epsilon_* \rho / t_{\rm ff}$, where $\dot\rho_*$ is the star formation rate density, $\epsilon_* = 0.02$ ([Kennicutt, 1998, Krumholz and Tan, 2007]) the constant star formation efficiency, and $t_{\rm ff}$ the local free-fall time of the gas. Star formation is allowed where the gas Hydrogen number density exceeds $n_0 = 0.1\,{\rm H\,cm^{-3}}$ according to a Poisson random process ([Rasera and Teyssier, 2006, Dubois and Teyssier, 2008]) with a stellar mass resolution of $M_{*,\rm res} = \rho_0 \Delta x^3 \simeq 2 \times 10^6\,{\rm M_\odot}$.

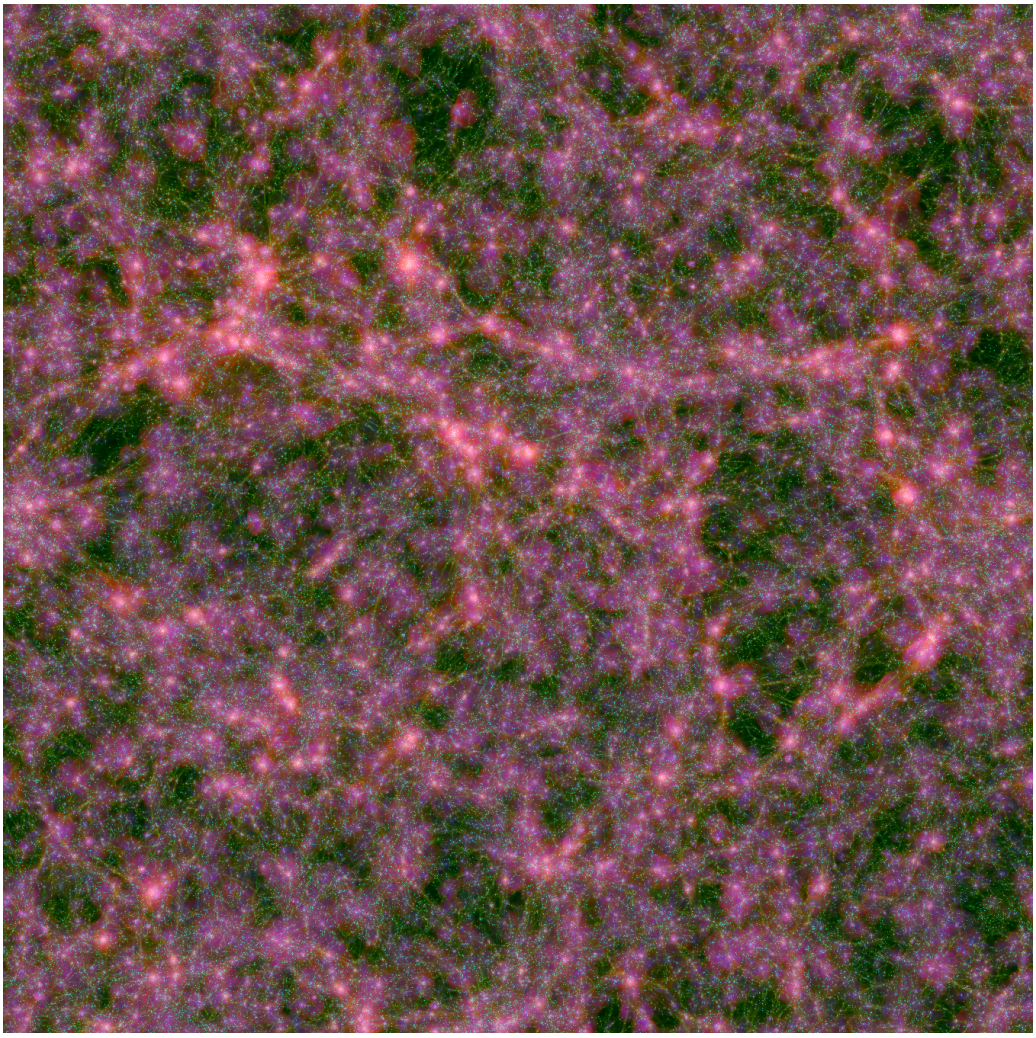


Figure 2.6: The Horizon-AGN simulation. A slice of the simulation at redshift z=1.2 is displayed with the gas density in green, the gas temperature in red, and the gas metallicity in blue.

The stellar feedback is modeled using a [Salpeter, 1955] initial mass function with a low-mass (high-mass) cut-off of $0.1\,\mathrm{M}_\odot$ ($100\,\mathrm{M}_\odot$). In particular, the mechanical energy from supernovae type II and stellar winds follows the prescription of STARBURST99 ([Leitherer et al., 1999, Leitherer et al., 2010]), and the frequency of type Ia supernovae explosions is taken from [Greggio and Renzini, 1983].

Active Galactic Nuclei (AGN) feedback is modeled according to [Dubois et al., 2012]. A Bondi-Hoyle-Lyttleton accretion rate onto Black Holes is used $\dot{M}_{\rm BH} = 4\pi\alpha G^2 M_{\rm BH}^2 \bar{\rho}/(\bar{c}_s^2 + \bar{u}^2)^{3/2}$, where $M_{\rm BH}$ is the BH mass, $\bar{\rho}$ is the average gas density, $\bar{c}_s$ is the average sound speed, $\bar{u}$ is the average gas velocity relative to the BH velocity, and α is a dimensionless boost factor with $\alpha = (\rho/\rho_0)^2$ when $\rho > \rho_0$ and $\alpha = 1$ otherwise ([Booth and Schaye, 2009]) in order to account for our inability to capture the colder and higher density regions of the inter-stellar medium. The effective accretion rate onto BHs is capped at the Eddington accretion rate: $\dot{M}_{\rm Edd} = 4\pi G M_{\rm BH} m_{\rm p}/(\epsilon_{\rm r}\sigma_{\rm T} c)$, where $\sigma_{\rm T}$ is the Thompson cross-section, c is the speed of light, $m_{\rm p}$ is the proton mass, and $\epsilon_{\rm r}$ is the radiative efficiency, assumed to be equal to $\epsilon_{\rm r} = 0.1$ for the [Shakura and Sunyaev, 1973] accretion onto a Schwarzschild BH. Two different modes of AGN feedback are accounted for, the *radio* mode operating when $\chi = \dot{M}_{\rm BH}/\dot{M}_{\rm Edd} < 0.01$ and the *quasar* mode active otherwise. More details are given in [Dubois et al., 2014].

Galaxies are identified with the AdaptaHOP finder ([Aubert et al., 2004]), which relies directly on the distribution of star particles to construct the catalogue of galaxies. 20 neighbours are used to compute the local density of each particle. A local threshold of $\rho_{\rm t} = 178$ times the average total matter density is applied to select relevant densities. Note that the galaxy population does not depend sensitively on the exact value chosen for this threshold. Our specific choice reflects the fact that the average density of galaxies located at the centre of galaxy clusters is comparable to that of their host. The force softening (minimum size below which substructures are treated as irrelevant) is ~ 10 kpc. Only galactic structures identified with more than 50 star particles are included in the mock catalogues. This enables a clear identification of galaxies, including those in the process of merging. A galaxy catalogues with $\sim 165\,000$ objects is produced at $z = 1.2$ with masses between 1.7×10^8 and $1.4 \times 10^{12}\,\mathrm{M}_\odot$. The galaxy stellar masses quoted in this thesis should be understood as the sum over all star particles that belong to a galaxy structure identified by AdaptaHOP.

Successes and limitations Even if dark matter only simulations give accurate predictions for dark matter halo statistics, the effect of baryonic physics must still be taken into account and can lead to important deviation from dark matter only predictions. For instance, the matter power spectrum seems to be affected by baryons on scales $k \gtrsim 1h/\mathrm{Mpc}$. Baryons tend to also bias the halo mass function and the cosmic shear signal. This means in particular that percent precision cosmology definitely needs theorists to take into account the effect of baryonic physics when comparison between theory and observations is made. Given the uncertainties in modelling galaxy formation, this raises the question of the validity of hydrodynamical simulations on those scales.

2.2 Statistical estimators

The statistical description of the large-scale structure is based on several major concepts that this section aims at presenting.

2.2.1 Poly-spectra

Given a (classical) scalar random field such as the density contrast $\delta(x)$, on can define the joint probability distribution function $\mathcal{P}(\delta_1, \ldots, \delta_n)$ so that $\mathcal{P}(\delta_1, \ldots, \delta_n)\mathrm{d}\delta_1 \ldots \mathrm{d}\delta_n$ is the probability of having $\delta(x_1)$ between δ_1 and $\delta_1 + \mathrm{d}\delta_1$... and $\delta(x_n)$ between δ_n and $\delta_n + \mathrm{d}\delta_n$. The expectation value of any function f of the density is then given by

$$f_0 = \langle f \rangle = \int f(\delta_i) P(\delta_i) \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_n.$$

The ergodic principle can be used to replace this average on different realisations of the Universe by a spatial average

$$f_0 = \langle f(\mathbf{x}) \rangle = \frac{1}{V} \int f(\mathbf{x}) \, \mathrm{d}^3\mathbf{x} \, .$$

Hence, the statistics of a cosmic random field δ can be described by its successive N-point correlation functions. The first non-trivial such estimator is the two-point correlation function $\xi(r) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle$ that only depends on the magnitude of the separation $\mathbf{r}$ if the field is statistically homogeneous and isotropic. These two assumptions are equivalent to assume translation and rotation invariance. In Fourier space, it reads

$$\begin{aligned} \langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle &= \delta_D(\mathbf{k} + \mathbf{k}') \int \xi(r) \exp(i\mathbf{k} \cdot \mathbf{r}) \, \mathrm{d}^3\mathbf{r}, & (2.23) \\ &= \delta_D(\mathbf{k} + \mathbf{k}') P(k), & (2.24) \end{aligned}$$

where $P(k)$ is the power spectrum of $\delta(\mathbf{x})$. In higher dimensions, the (connected) correlation function between n planar waves define the successive poly-spectra P_n

$$\langle \delta(\mathbf{k}_1) \ldots \delta(\mathbf{k}_n) \rangle_c = \delta_D(\mathbf{k}_1 + \cdots + \mathbf{k}_n) P_n(\mathbf{k}_1, \ldots, \mathbf{k}_n).$$

For Gaussian random fields (such as the initial fields we are considering in this manuscript), Wick's theorem simply reads

$$\begin{aligned} \langle \delta(\mathbf{k}_1) \ldots \delta(\mathbf{k}_{2n+1}) \rangle &= 0, \\ \langle \delta(\mathbf{k}_1) \ldots \delta(\mathbf{k}_{2n}) \rangle &= \sum_k \prod_{(i,j) \in \Sigma_k} \langle \delta(\mathbf{k}_i) \delta(\mathbf{k}_j) \rangle, \end{aligned}$$

where Σ enumerates all the different possibilities of gathering the fields by pairs. This property means that Gaussian fields are fully described by their two-point correlation function or equivalently their power spectrum.

2.2.1.1 The initial power spectrum

Primordial density inhomogeneities are assumed to form a stochastic field as the result of quantum fluctuations in the primordial Universe. The observed near-Gaussianity of the temperature fluctuations in the cosmic microwave background allows one to assume that the statistics of the initial density fluctuations are Gaussian so that the initial joint probability distribution function (PDF) of $X = (\delta_1, \ldots, \delta_n)$ is

$$\mathcal{P}(\delta_1, \ldots, \delta_n) = \frac{1}{\sqrt{\det |2\pi \mathbf{C}|}} \exp\left[-\frac{1}{2} \mathbf{X}^t \cdot \mathbf{C}^{-1} \cdot \mathbf{X} \right], \quad (2.25)$$

where $\mathbf{C}$ is the covariance matrix describing the correlations of the fluctuation field $C_{ij} = \langle \delta_i \delta_j \rangle$. The power-spectrum of primordial fluctuations has been shown by cosmic microwave background

experiments to be very close to scale-invariant i.e $P(k) \propto k^n$, $n \approx 1$. This power spectrum is also known as the Harrison-Zel'dovich spectrum and is such that perturbations in the gravitational potential are scale-independent. Indeed, the Poisson equation implies that the power spectrum, $P_\Phi(k)$, of the gravitational potential goes like $k^{n-4} \approx k^{-3}$ in which case, the amplitude of perturbations

$$\sigma(\Phi)^2 = \frac{1}{(2\pi)^3}\int \mathrm{d}^3\,\mathbf{k}\,P_\Phi(k) = \frac{1}{2\pi^2}\int \mathrm{d}\log k\,k^3 P_\Phi(k) \tag{2.26}$$

is independent of scale $\mathrm{d}\sigma^2(\Phi)/\mathrm{d}\log k = \mathrm{cst}$. This scale-invariant property is often presented as a prediction of the standard inflationary scenario. Modes that have been stretched to super-horizon scales by inflation then start growing as soon as they re-enter the horizon. The evolution of perturbation in the radiation-dominated era can be encoded in a transfer function T that relates the primordial nearly scale-invariant power spectrum to the power spectrum after the matter-radiation equivalence. This initial power spectrum (i.e at the beginning of the matter-dominated epoch) will be called $P_0(k)$ in what follows and is given by

$$P_0(k,z) = A(z)k^n T^2(k,z), \tag{2.27}$$

where the normalisation $A(z)$ is determined observationally. The transfer function that encodes the microphysics of the recombination is well-approximated by [Bardeen et al., 1986]

$$T_{\mathrm{CDM}} = \frac{\ln(1+2.34q)}{2.34q}(1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4)^{-1/4} \tag{2.28}$$

where $q = k/\Gamma h/\mathrm{Mpc}$ and $\Gamma = \Omega_0 h \exp\left[-\Omega_b(1+\sqrt{2h}/\Omega_0)\right]$. On very large scales, the power spectrum is scale-invariant as set by inflation (fluctuations are frozen). On the other hand, this power spectrum on small scales goes like k^{-3}. The turnover point occurs at the comoving Hubble radius around the matter-radiation equivalence $r_H(t_{\mathrm{eq}}) = 13.7/(\Omega_{m,0}h^2)$ Mpc.

2.2.1.2 Smoothing

The subsequent gravitational evolution is then fully described by the Newtonian formalism described in section 2.1.2. The idea is now to investigate how the initial statistics seen in the CMB evolve under gravity and how non-Gaussianities arise and skew the probability density function. In order to compare with observations, it is necessary to introduce filter functions that allow one to go from a point-like distribution of galaxies to a density field smoothed on a given scale. The two most common smoothing functions are the Gaussian filter

$$W_G(\mathbf{x},R) = \frac{1}{(2\pi)^{3/2}R^3}\exp\left(\frac{-|\mathbf{x}|^2}{2R^2}\right), \tag{2.29}$$

and the top-hat filter which averages the matter content in hard spheres of radius R

$$W_{\mathrm{TH}}(\mathbf{x},R) = \frac{3}{4\pi R^3}\Theta_H(R-|\mathbf{x}|)\,. \tag{2.30}$$

The normalisation of those filter functions is a convention. It is set to one here

$$\int \mathrm{d}^3\mathbf{x}\,W(\mathbf{x},R) = 1\,. \tag{2.31}$$

In Fourier space, the filters read

$$W_G(k,R) = \frac{1}{(2\pi)^{3/2}}\int \mathrm{d}^3\mathbf{x}\,\exp\left(-i\mathbf{k}\cdot\mathbf{x}\right)W_G(\mathbf{x},R) = \frac{1}{(2\pi)^{3/2}}\exp\left(\frac{-k^2R^2}{2}\right)\,, \quad (2.32)$$

$$W_{\rm TH}(k,R) = \frac{1}{(2\pi)^{3/2}}\frac{3}{(kR)^2}\left(\frac{\sin(kR)}{(kR)}-\cos(kR)\right)\,, \quad (2.33)$$

where k is the norm of the wavevector $k=|\mathbf{k}|$.

Note that in order to study the same level of non-linearity, one needs to use either one filter or the other with a different smoothing. To be more precise, let us compute the variance of fluctuations

$$\sigma^2(R) = \frac{1}{(2\pi)^3}\int 4\pi k^2\mathrm{d}k\,P(k)W^2(k,R)\,. \quad (2.34)$$

Then the link between top-hat and Gaussian smoothing lengths is implicitly given by $\sigma^2(R_{\rm TH}) = \sigma^2(R_G)$. For a power-law power spectrum $P(k)\propto k^n$, we get $R_{\rm TH}=\alpha(n)R$ where

$$\alpha^{n+3}(n) = 9\times 2^{2-n}\frac{\Gamma(n-1)\sin(n\frac{\pi}{2})}{(n-3)\Gamma\left(\frac{n+1}{2}\right)}. \quad (2.35)$$

For instance, $\alpha(n=0)=(6\sqrt{\pi})^{1/3}\approx 2.2$.

2.2.1.3 Non-linear evolution of the power spectrum

The effect of the dynamical evolution of the density field is already seen in the power spectrum. Its non-linear evolution can be studied through perturbation theory in the weakly non-linear regime as described in section 2.1.2.4 or numerical simulations. At linear order, the global magnitude of the power spectrum grows like $D_+(t)^2$ as shown in section 2.1.2.2. Next order corrections can be computed through perturbation theory techniques until two loops. In [Taruya et al., 2012], we showed how to predict the density power spectrum to two-loop order using the RegPT approach. Figure 2.7 compares this prediction to the result of dark matter simulations. Such a construction gives robust and accurate predictions for both the density power spectrum and the correlation function at percent-level in the weakly non-linear regime. At decreasing redshift, the clustering non-linearity develops and the applicable range of PT calculations inevitably becomes narrower. However, compared to the standard PT predictions, the RegPT result can reproduce the N-body trend with even wider range. Note that the impact of non-linear clustering on the baryon acoustic peak of the two-point correlation function is significant (right panel): the peak position becomes slightly shifted to a smaller scale, and structure of the peak tends to be smeared as the redshift decreases. The RegPT (similarly to other improved PT treatments) calculation can describe not only the behavior around the baryon acoustic peak but also the small-scale behavior of the correlation function. Although the RegPT prediction eventually deviates from simulations at small scales and the result at $z=0.35$ indeed manifests the discrepancy below $r\approx 30\,h^{-1}$Mpc, the actual range of agreement between RegPT and N-body results is wider than what is naively expected from the power spectrum results. In fact, it has been recently advocated by several authors that with the improved PT treatment, the one-loop calculation is sufficient to accurately describe the two-point correlation function.

All the cosmological information is not contained in the power spectrum because the non-linear evolution also induces non-Gaussianities that we propose to describe in the following section.

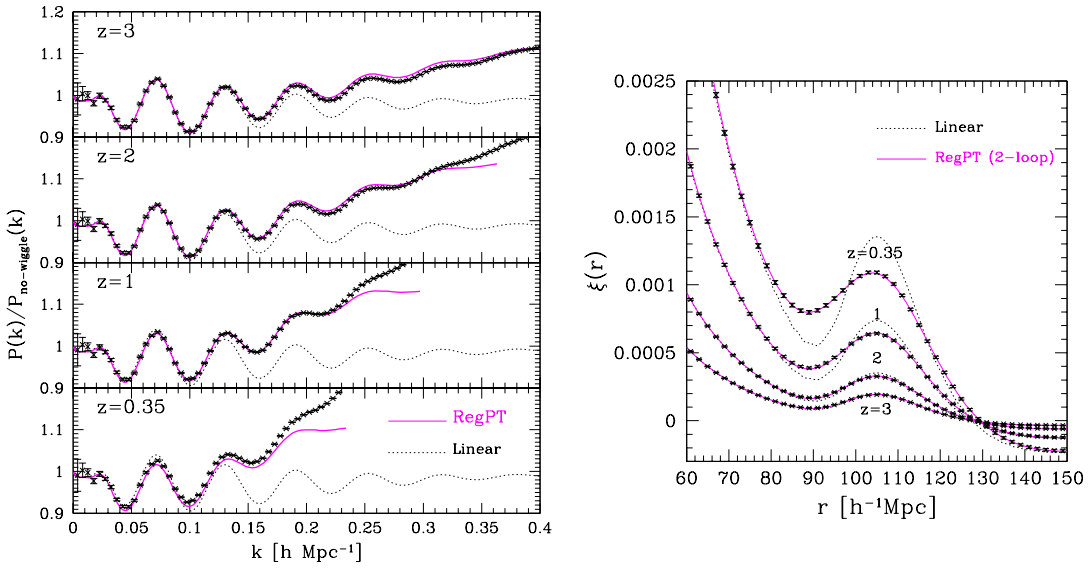


Figure 2.7: Comparison of density power spectrum (left panel) and two-point correlation function (right panel) between N-body simulations and RegPT calculations. Note that left panel shows the ratio of the power spectrum to the smooth linear spectrum, $P_{\text{no-wiggle}}(k)$, calculated from the no-wiggle formula of the linear transfer function. For reference, the linear prediction is also displayed (dotted line). These figures are reproduced from [Taruya et al., 2012].

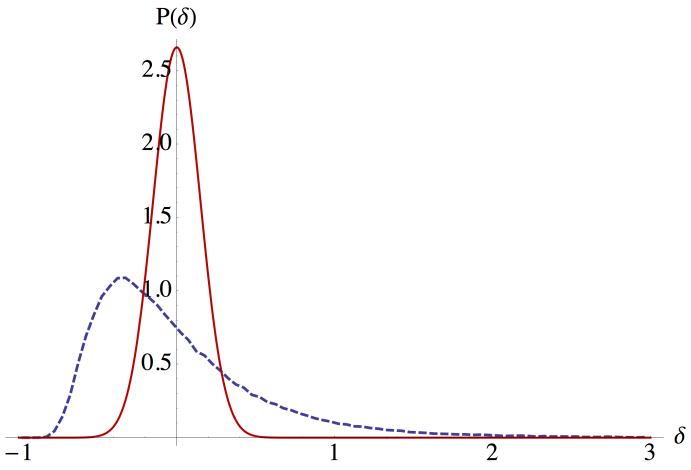


Figure 2.8: Non-linear evolution of the density PDF from an initial Gaussian distribution (red) to a skewed PDF (blue) as measured in a numerical simulation at redshift $z = 0.65$ for a top-hat smoothing on scale $R_{\rm TH} = 10$ Mpc/h.

2.2.1.4 Non-Gaussianities

When structure grow, peaks are becoming more and more dense in comoving space and voids are expanding. This can be seen in the density distribution with a long high-density tail and a pronounced peak at low-density (see figure 2.8). As a consequence, on top of the growth of its variance, the density PDF is becoming increasingly non-Gaussian and in particular asymmetric with cosmic time. The skewness (that measures the asymetry of the PDF) $\langle\delta^3\rangle$ and other higher order moments become non-zero and grow.

Let us compute the skewness at leading order in perturbation theory. If initial conditions are Gaussian ([Peebles, 1980])

$$\langle\delta^3\rangle = 3\left\langle\left(\delta^{(1)}\right)^2\delta^{(2)}\right\rangle + \mathcal{O}(\sigma^2) = \frac{34}{7}\langle\delta^2\rangle^2 + \mathcal{O}(\sigma^2)\,. \tag{2.36}$$

This result neglects the effect of smoothing. Adding a filter function increases the complexity of the angular part of the integration. However, in the case of a top-hat filter in real space (equation 2.30), the properties of Bessel functions allow to analytically compute the skewness so that ([Bernardeau, 1994a])

$$\langle\delta^3(R)\rangle = \left(\frac{34}{7} + \frac{\mathrm{d}\log\sigma^2(R)}{\mathrm{d}\log R}\right)\langle\delta^2\rangle^2 + \mathcal{O}(\sigma^2)\,. \tag{2.37}$$

As the skewness scales like σ^4 at tree order (i.e on large scales), it is common to define the S_3 parameter as

$$S_3 = \frac{\langle\delta^3(R)\rangle}{\langle\delta^2(R)\rangle^2} = \frac{34}{7} + \frac{\mathrm{d}\log\sigma^2(R)}{\mathrm{d}\log R} + \mathcal{O}(\sigma^2)\,. \tag{2.38}$$

Similarly, one can define the successive S_n factors as the ratio of the n-th order cumulant to some power of the variance

$$S_n = \frac{\langle\delta^n(R)\rangle_c}{\langle\delta^2(R)\rangle^{n-1}}\,. \tag{2.39}$$

Those factors are pure numbers on large-scales (see [Fry, 1984, Bernardeau, 1994a, Lokas et al., 1995] for S_4, the kurtosis). The reason of this property is twofold: Gaussian initial conditions and quadratic PT kernels. It can be easily understood from the diagrammatic representation of a cumulant of order n as described in [Bernardeau et al., 2002].

The effect of the S_n factors on the density PDF can be appreciated via the so-called Edgeworth expansion of the PDF in powers of σ ([Scherrer and Bertschinger, 1991, Bernardeau and Kofman, 1995, Juszkiewicz et al., 1995])

$$\mathcal{P}(\delta) = G(\delta) \times \left[1 + \sigma \frac{S_3}{3!} H_3\left(\frac{\delta}{\sigma}\right) + \sigma^2 \left(\frac{S_4}{4!} H_4\left(\frac{\delta}{\sigma}\right) + \frac{1}{2}\left(\frac{S_3}{3!}\right)^2 H_6\left(\frac{\delta}{\sigma}\right)\right) + \cdots\right], \tag{2.40}$$

where $G(\delta) = \exp\left(-\delta^2/(2\sigma^2)\right)/\sqrt{2\pi\sigma^2}$ is the Gaussian distribution that matches the mean and variance of $\mathcal{P}$ and S_n can be viewed as non-Gaussian perturbations of the PDF around a Gaussian kernel. This expansion can be derived from the cumulant generating function

$$\varphi(\lambda) = \sum_{p\geq 2} (-1)^{p-1} \frac{S_p}{p!} \tag{2.41}$$

that is related to the density PDF through

$$\mathcal{P}(\delta) = \frac{1}{2i\pi\sigma^2} \int_{-i\infty}^{i\infty} \mathrm{d}\lambda \exp\left(-\frac{\varphi(\lambda)}{\sigma^2} + \frac{\delta\lambda}{\sigma^2}\right). \tag{2.42}$$

At fixed S_p, this PDF can be computed when σ and λ are small leading to the Edgeworth expansion.

This Edgeworth series is a re-ordering of a Gram-Charlier expansion ([Cramér, 1946, Kendall and Stuart, 1958, Chambers, 1967, Juszkiewicz et al., 1995, Amendola, 1996, Blinnikov and Moessner, 1998]) which allows one to expand a PDF around a (e.g Gaussian) distribution using the corresponding basis of orthogonal polynomials (here Hermite polynomials)

$$\mathcal{P}(x) = G(x) \times \left[1 + \sum_{n=3}^{\infty} \frac{\langle x^n \rangle_{\mathrm{GC}}}{n!} H_n(x)\right], \tag{2.43}$$

where for the sake of simplicity $x = \delta/\sigma$ and the Gram-Charlier coefficients can easily be derived from the orthogonality property of Hermite polynomials

$$\langle x^n \rangle_{\mathrm{GC}} = \langle H_n(x) \rangle . \tag{2.44}$$

Let us recall that Hermite polynomials are defined such that

$$\frac{\mathrm{d}^n \exp(-x^2/2)}{\mathrm{d}x^n} = (-1)^n H_n(x) \exp(-x^2/2), \tag{2.45}$$

so that for instance $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - x$, etc.

Gram-Charlier expansions suffer from a lack of convergence. This is solved by the Edgeworth series which re-orders in powers of the variance. However, it has to be noted that both approaches are valid only in the weakly non-linear regime and are not positive-definite. Nevertheless, it has been realised that the count-in-cell formalism allows one to predict the density PDF in the *mildly* non-linear regime using the spherical collapse model. Those ideas will be developed in section 5.1.2. Note also that the PDF expansions presented in this section will be an important

tool to predict topological and geometrical estimators in the weakly non-linear regime (see section 5.1.1).

The non-Gaussian evolution of the density field is not only encoded in its one-point PDF (or in the one-point cumulants S_n) but in the full hierarchy of N-point correlation functions. Beyond the two-point correlation function (or its Fourier counterpart the power spectrum) that is sufficient to fully describe a Gaussian field, in the non-Gaussian regime one is led to also consider the three-point correlation function (or the bispectrum), the four-point correlation function (or the trispectrum), etc that are increasingly difficult to measure and predict. One important question that arises at this point is the building of "optimal" observables that carry as much cosmological information as possible. One idea is for instance to focus on topological estimators (see the following section).

2.2.2 Topology of the density field

With upcoming high-precision surveys, it has become necessary to revisit alternative tools to investigate the statistics of random cosmological fields so as to handle observables with different sensitivity. Minkowski functionals ([Mecke et al., 1994]) have been being actively used ([Gott et al., 1987, Weinberg et al., 1987, Melott et al., 1988, Gott et al., 1989, Hikage et al., 2002, Hikage et al., 2003, Park et al., 2005, Gott et al., 2007, Planck Collaboration et al., 2014d]) as an alternative to the usual direct measurements of higher-order moments and N-point correlation functions ([Scoccimarro et al., 1998, Percival et al., 2007, Gaztañaga et al., 2009, Nishimichi et al., 2010, amongst many other studies]). These functionals describe the topological properties of random fields. As such, they will present different biases and might be more robust, e.g. with regards to rare events.

2.2.2.1 Definitions

In order to study the geometry of a field, one can focus on its set of excursions – the volume above a given threshold – or equivalently on its sets of isocontours (surface of constant value of the field and therefore boundary of an excursion). Figure 2.9 shows a typical excursion. Then the study of their topology leads to the so-called Minkowski functionals that we propose to study here.

The topology of a 2D manifold is characterized via the so-called mathematical genus g which is its number of handles (zero for the surface of a sphere, one for a torus, two for a double torus, etc). This quantity is trivially invariant under continuous deformation. In arbitrary dimension, this notion is replaced by the Euler-Poincaré characteristic χ defined as the alternating sum of Betti numbers

$$\chi = \sum_i (-1)^i b_i, \tag{2.46}$$

where b_i is the rank of the i-th singular homology group H_i (which counts the holes of dimension i). For the torus, $b_0 = 1$, $b_1 = 2$, $b_2 = 1$ so that the Euler-Poincaré characteristic of a torus is $1 - 2 + 1 = 0$. For 2D manifold, χ is related to the genus by $\chi = 2 - 2g$ and the Gauss-Bonnet theorem states that the 2D Euler-Poincaré characteristic is nothing but the integral of the Gaussian curvature K allowing to relate the topology to the geometry of the surface

$$\chi = \frac{1}{2\pi}\left(\int_S K \mathrm{d}A + \int_{\partial S} k \mathrm{d}S\right), \tag{2.47}$$

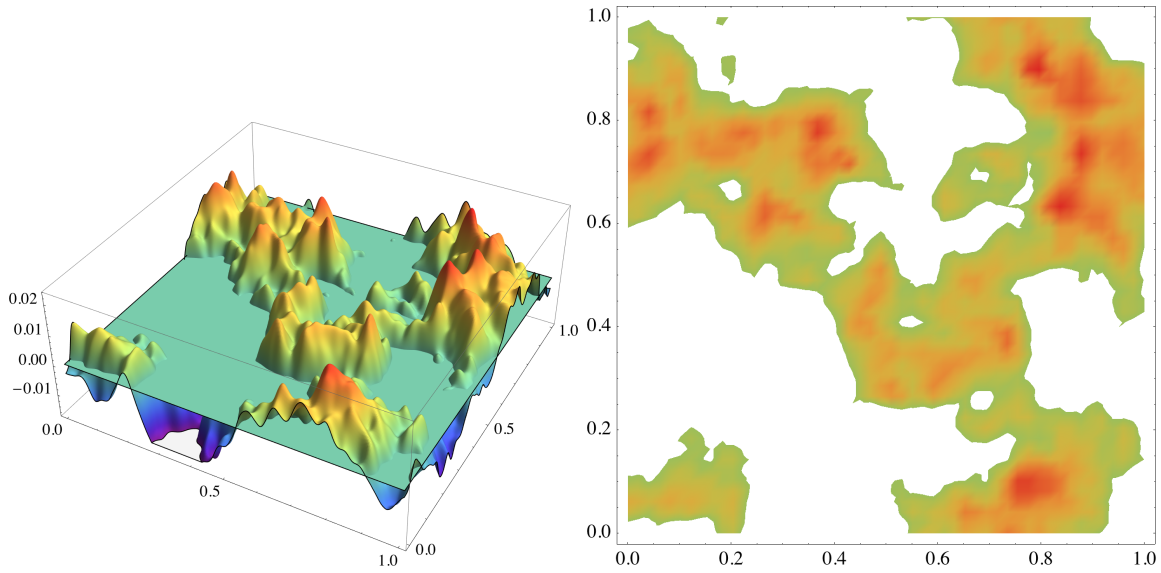


Figure 2.9: Excursion above the threshold $f = 0$ of the 2D Gaussian field f.

where k is the curvature of the boundary of the surface S. Note that in cosmology, we are interested in the topological features of either the excursions or the isocontours. Their Euler characteristics are related by

$$\chi(\partial M) = (1 + (-1)^{d-1})\chi(M), \tag{2.48}$$

where M is a d-dimensional manifold (e.g an excursion) and ∂M its boundary (e.g an isocontour).

A finite number of invariants is sufficient to characterize the topology of random fields, they are known as Minkowski functionals (and are functions of the threshold ν). More precisely, they are the only morphological descriptors in Integral Geometry that respect motion-invariance, conditional continuity and additivity ([Hadwiger, 1957]). As a result, they form a robust and meaningful set of observables that characterizes the excursion sets or equivalently the isocontours of the field. In d dimensions, there are $d + 1$ such functionals (4 in 3D and 3 in 2D), namely in 3D: the encompassed volume, f_V, the surface area, $\mathcal{N}_3$, the integral mean curvature and the integral Gaussian curvature which is closely related to the Euler-Poincaré characteristic, χ (see also figure 2.10). For random fields these functionals are understood as densities, i.e quantities per unit volume of space. Especially when studying anisotropic fields, complimentary information can be obtained by using Minkowski functionals for the field obtained on lower dimensional sections of the 3D field. For example, in addition to 3D isocontour area statistics, one can introduce the length of 2D isocontours on a planar sections $\mathcal{N}_2$, and contour crossings by a line through 3D space, $\mathcal{N}_1$. These statistics for cosmology were first introduced by [Ryden, 1988, Ryden et al., 1989]. In the isotropic limit, they are trivially related: $2\mathcal{N}_1 = 4\mathcal{N}_2/\pi = \mathcal{N}_3$; but this relation does not hold anymore for an anisotropic field as it will be shown in section 5.1.1.1. Similarly, in addition to the full Euler characteristic χ_{3D} of 3D excursion sets, we shall consider the 2D Euler characteristic, χ_{2D}, on planar sections through the field.

Among all Minkowski functionals, the beauty of the Euler characteristic is that it can be related to the critical points via Morse theory. Indeed, if the field is smooth enough (in mathematical terms, if it is a Morse function), the Euler characteristic is nothing but the alternating sum of

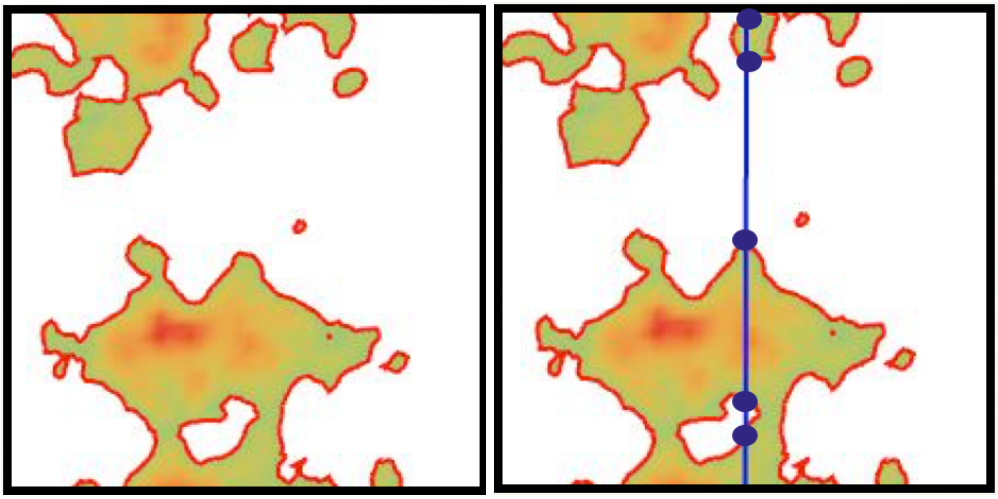


Figure 2.10: In two dimensions, there are three Minkowski functionals: the Euler characteristics (related to the genus) which counts the number of islands minus the number of holes, the length of the isocontour (left panel, red line) and the volume above the threshold. Contour crossing statistics are illustrated on the right panel.

extrema

$$\chi_{2\mathrm{D}}(\nu) = n_{\min}(\nu) - n_{\mathrm{saddle}}(\nu) + n_{\max}(\nu)\,, \tag{2.49}$$

$$\chi_{3\mathrm{D}}(\nu) = n_{\min}(\nu) - n_{\mathrm{sadw}}(\nu) + n_{\mathrm{sadf}}(\nu) - n_{\max}(\nu)\,. \tag{2.50}$$

This equivalence between Euler characteristic and critical points is illustrated in figure 2.11 in the simple case of a one-dimensional field. It is clear on this figure that maxima indeed create connected components and minima destroy holes and merge components together.

Similarly to the Euler characteristic, other Minkowski functionals can be expressed as averages over the joint probability density function (JPDF) of the field and its derivatives. In the following, let us call x the field under consideration and, without loss of generality, assume that it has zero mean. In cosmological applications, this field, for instance, can be the 3D density contrast.

2.2.2.2 Peak theory

For the sake of simplicity, let us first explain how to use the JPDF to study the number count of peaks in a random field x. The formalism of cosmological density peaks, which builds on the Kac-Rice formula [Kac, 1943, Rice, 1945] was laid down in [Bardeen et al., 1986]. Following [Pogosyan et al., 2009b], for a given field ρ, we define the moments

$$\sigma_0{}^2 = \langle \rho^2 \rangle, \qquad \sigma_1{}^2 = \langle (\nabla\rho)^2 \rangle, \qquad \sigma_2{}^2 = \langle (\Delta\rho)^2 \rangle. \tag{2.51}$$

Combining these moments, we can build two characteristic lengths $R_0 = \sigma_0/\sigma_1$ and $R_* = \sigma_1/\sigma_2$, as well as the spectral parameter

$$\gamma = \frac{\sigma_1{}^2}{\sigma_0\sigma_2}. \tag{2.52}$$

We choose to normalise the field and its derivatives to have unit variances:

$$x = \frac{1}{\sigma_0}\rho, \qquad x_i = \frac{1}{\sigma_1}\nabla_i\rho, \qquad x_{ij} = \frac{1}{\sigma_2}\nabla_i\nabla_j\rho. \tag{2.53}$$

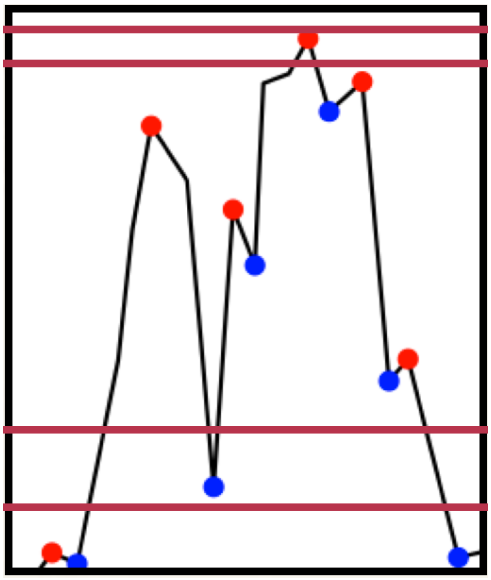


Figure 2.11: The alternating sum of extrema is nothing but the Euler characteristics.

In terms of these variables, the Gaussian joint PDF for the field and its derivatives, $\mathbf{X} = \{x, x_{ij}, x_i\}$ is the multivariate Normal distribution

$$\mathcal{P}(\mathbf{X}) = \frac{\exp\left(-\frac{1}{2}\mathbf{X}^{\mathrm{T}} \cdot \mathbf{C_0}^{-1} \cdot \mathbf{X}\right)}{\det|\mathbf{C_0}|^{1/2}\,(2\pi)^{(\mathrm{d}+1)(\mathrm{d}+2)/2}}\,, \tag{2.54}$$

where $\mathbf{C}_0 \equiv \langle \mathbf{X} \cdot \mathbf{X}^{\mathrm{T}} \rangle$ is the covariance matrix. Eq.(2.54) is sufficient to compute the expectation of any quantity involving the fields and its derivatives up to second order, in particular the expected number density of peaks. Indeed the number density of peaks can be written as a sum of Dirac delta functions centered on the locations $\mathbf{r_i}$ of the peaks

$$n_{\max}(\mathbf{r}) = \sum_{1 \geq i \geq N_{\max}} \delta_{\mathrm{D}}(\mathbf{r} - \mathbf{r_i})\,. \tag{2.55}$$

Let us now write a Taylor expansion of the density gradient around a peak

$$\nabla x(\mathbf{r}) = 0 + \mathbf{H}(\mathbf{r_i}) \cdot (\mathbf{r} - \mathbf{r_i})\,. \tag{2.56}$$

where $\mathbf{H}$ is the tensor of second derivatives (Hessian matrix) of the field x. The properties of the Dirac delta function then yields

$$|\det \mathbf{H}|\delta_D(\nabla x) = \sum_{1 \geq i \geq N_{\mathrm{crit}}} \delta_{\mathrm{D}}(\mathbf{r} - \mathbf{r_i})\,, \tag{2.57}$$

where the sum runs on all the critical points ($\delta_D(\nabla x) = 0$) of the field.

Combining equations 2.55 and 2.57, one eventually gets the number density of critical points

$$\langle n_{\mathrm{crit}}(\mathbf{r}) \rangle = \langle |\det \mathbf{H}|\delta_D(\nabla x) \rangle\ , \tag{2.58}$$

and the number density of peaks by adding a constraint on the sign of the eigenvalues λ_k of the Hessian $\mathbf{H}$

$$\langle n_{\max}(\mathbf{r}) \rangle = \left\langle |\det \mathbf{H}|\delta_D(\nabla x) \prod_k \Theta(-\lambda_k) \right\rangle\ , \tag{2.59}$$

where averages are turned from spatial averages to ensemble averages (by ergodicity)

$$\langle n_{\rm max}\rangle = \left\langle |\det \mathbf{H}|\delta_D(\nabla x)\prod_k \Theta(-\lambda_k)\right\rangle \tag{2.60}$$

$$= \int {\rm d}\mathbf{X}\,\mathcal{P}(\mathbf{X})|\det \mathbf{H}|\delta_D(\nabla x)\prod_k \Theta(-\lambda_k) \tag{2.61}$$

$$= \left(\frac{\sigma_2}{\sigma_1}\right)^d \int {\rm d}\mathbf{X}\,\mathcal{P}(\mathbf{X})|\det x_{ij}\prod_k \Theta(-\lambda_k)|\delta_D(x_k)\,. \tag{2.62}$$

This integral can be exactly calculated for a Gaussian JPDF in two and three dimensions so that the total number density of peaks (and similarly saddles and minima) is given in two dimensions by

$$\langle n_{\rm max}\rangle = \langle n_{\rm min}\rangle = \frac{1}{8\sqrt{3}\pi R_\star^2} \tag{2.63}$$

$$\langle n_{\rm sad}\rangle = \frac{1}{4\sqrt{3}\pi R_\star^2} \tag{2.64}$$

and in three dimensions by

$$\langle n_{\rm max}\rangle = \langle n_{\rm min}\rangle = \frac{29\sqrt{15}-18\sqrt{10}}{1800\pi^2 R_\star^3} \tag{2.65}$$

$$\langle n_{\rm sadf}\rangle = \langle n_{\rm sadw}\rangle = \frac{29\sqrt{15}+18\sqrt{10}}{1800\pi^2 R_\star^3}, \tag{2.66}$$

where filament-type saddle points (two negative eigenvalues) are dissociated from wall-type saddle points (one negative eigenvalue).

The same formalism can be used to compute the number density of peaks (saddles, minima) of a given height ν. While in two dimensions the result is still analytical, this is not the case anymore in three dimensions but integrals can be computed from a numerical point of view. Peak counts can also be predicted for weakly non-Gaussian fields using a Gram-Charlier expansion of the JPDF as shown in [Gay et al., 2012].

2.2.2.3 Predicting Minkowski functionals

Minkowski functionals can be predicted from the knowledge of the JPDF, in the same way as peak counts. Collecting the well-known results from an extensive literature (e.g. [Rice, 1944, Rice, 1945, Ryden, 1988, Matsubara, 1996]) in a compact form, we have for the first two Minkowski functionals

$$f_V(\nu) = \langle\Theta(x-\nu)\rangle\,, \tag{2.67}$$

$$\mathcal{N}_3(\nu) = \langle|\nabla x|\delta_{\rm D}(x-\nu)\rangle\,, \tag{2.68}$$

$$\mathcal{N}_2(\nu) = \langle|\nabla_{\mathcal{S}}\, x|\delta_{\rm D}(x-\nu)\rangle\,, \tag{2.69}$$

$$\mathcal{N}_1(\nu) = \langle|\nabla_{\mathcal{L}}\, x|\delta_{\rm D}(x-\nu)\rangle\,, \tag{2.70}$$

where the $\delta_{\rm D}$-function in the statistical averaging signifies evaluation at the given threshold $x=\nu$, while the step function reflects the cumulative averaging over the values above the

threshold $x \geq \nu$. We see that the family of threshold-crossing statistics is given by the average gradient of the field for $\mathcal{N}_3$ or its restriction to a plane $\mathcal{S}$ or line $\mathcal{L}$ for $\mathcal{N}_2$ and $\mathcal{N}_1$, respectively.

The average of the Gaussian curvature on the isosurface is, via the Gauss-Bonnet theorem, its topological Euler characteristic χ, which thus can be expressed directly as ([Hamilton et al., 1986, Matsubara, 1996])

$$\chi(\nu) = \left\langle \delta_{\rm D}(x-\nu)\delta_{\rm D}(\nabla_1 x)\delta_{\rm D}(\nabla_2 x)\nabla_3 x|(\nabla_1\nabla_1 x\nabla_2\nabla_2 x - (\nabla_1\nabla_2 x)^2)\right\rangle . \tag{2.71}$$

In this thesis, we will use the Euler characteristic, $\chi_{\rm 3D}$, of the excursion set encompassed by the isosurface (which in 3D is just one half of the Euler charasteristic of the isosurface itself, and is equal to minus the genus for the definitions used in cosmology, see detailed discussions for such conventions in [Gay et al., 2012]). Being the alternating sum of Betti numbers, $\chi_{\rm 3D}$ is related via Morse theory (e.g. [Jost, 2008]) to the alternating sum of the number of critical points in the excursion volume. For a random field ([Doroshkevich, 1970, Adler, 1981, Bardeen et al., 1986]) [3]

$$\chi_{\rm 3D}(\nu) = -\left\langle \det(\nabla_i\nabla_j x)\,\delta_{\rm D}(\nabla x)\Theta(x-\nu)\right\rangle \ , \quad i,j \in \{1,2,3\} . \tag{2.72}$$

The Euler characteristic of the excursion sets of a 2D field (in particular, 2D slices of a 3D random field) is given by a similar expression ([Adler, 1981, Bond and Efstathiou, 1987, Coles, 1988, Melott et al., 1989, Gott et al., 1990])

$$\chi_{\rm 2D}(\nu) = \left\langle \det(\nabla_i\nabla_j x)\,\delta_{\rm D}(\nabla x)\Theta(x-\nu)\right\rangle \ , \quad i,j \in \{1,2\} . \tag{2.73}$$

Within the same formalism and as shown in section 2.2.2.2, it is easy to compute the critical points counts ([Adler, 1981, Bardeen et al., 1986]). Equation 2.72 leads to a cumulative counting above a given threshold for maxima, two type (filamentary and wall-like) saddle points and minima

$$n_{\rm max,3D}(\nu) = -\left\langle \det(\nabla_i\nabla_j x)\,\delta_{\rm D}(\nabla x)\Theta(-\lambda_1)\Theta(x-\nu)\right\rangle , \tag{2.74}$$

$$n_{\rm sadf,3D}(\nu) = +\left\langle \det(\nabla_i\nabla_j x)\,\delta_{\rm D}(\nabla x)\Theta(\lambda_1)\Theta(-\lambda_2)\Theta(x-\nu)\right\rangle , \tag{2.75}$$

$$n_{\rm sadw,3D}(\nu) = -\left\langle \det(\nabla_i\nabla_j x)\,\delta_{\rm D}(\nabla x)\Theta(\lambda_2)\Theta(-\lambda_3)\Theta(x-\nu)\right\rangle , \tag{2.76}$$

$$n_{\rm min,3D}(\nu) = +\left\langle \det(\nabla_i\nabla_j x)\,\delta_{\rm D}(\nabla x)\Theta(\lambda_3)\Theta(x-\nu)\right\rangle , \tag{2.77}$$

where averaging conditions are set by the signs of sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of the Hessian matrix of the field. Taking alternating sum eliminates the constraints on signs of eigenvalue, leading to the $\chi_{\rm 3D}$ statistics. Similar expressions as equations 2.74-2.77 apply for 2D extrema.

Extrema counts provide us with information on peaks (dense regions), minima (under-dense regions), and saddle points. In some applications there is symmetry between extrema (e.g. in CMB studies minima and maxima of the temperature field are equivalent); in others, they describe very different structures, e.g. in large-scale structure dense peaks correspond to gravitationally collapsing objects like galactic or cluster haloes while minima seed the regions devoid of structures. Saddle-type extrema are also interesting in their own right, being related to the underlying filamentary structures (bridges connecting peaks through saddles), which in turn can also be characterized by the skeleton of the cosmic web (see section 2.3). A particular advantage of the described geometrical and topological statistical estimates is that they are invariant under monotonic transformation of the underlying field $x \to f(x)$, provided one maps the threshold correspondingly $\nu \to f(\nu)$. For cosmological data this means that these statistics are formally invariant with respect to any monotonic local bias between the galaxy and matter distributions.

[3] Note that this expression comes from $\delta_{\rm D}(\nabla x) = \sum_{x_0|\nabla x_0=0} \delta_{\rm D}(x-x_0)/|\det(\nabla_i\nabla_j x)|$ and the absolute value of the Hessian can be dropped because we are interested in the *alternating* sum of critical points.

For Gaussian isotropic fields, the theoretical prediction of Minkowski functionals has long been known ([Doroshkevich, 1970, Adler, 1981, Bardeen et al., 1986, Hamilton et al., 1986, Bond and Efstathiou, 1987, Ryden, 1988, Ryden et al., 1989]). For instance, the Euler characteristic in three dimensions is simply given by

$$\chi_{3\mathrm{D}}(\nu) = \frac{\sqrt{3}}{36\pi^2}\frac{\gamma^3}{R_\star^3}H_2(\nu)\exp\left(-\nu^2/2\right) . \tag{2.78}$$

For mildly non-Gaussian fields, their expressions have also been computed more recently. The idea is to expand the JPDF of the field and its derivatives using either the first and second non-Gaussian corrections of a multivariate Edgeworth expansion ([Matsubara, 1994, Pogosyan et al., 2009a, Matsubara, 2010]) or the fully non-Gaussian Gram-Charlier expansion ([Gay et al., 2012]). However, one major assumption in these results has been the isotropy of the underlying field. In section 5.1.1.1, we extend the theory to partially anisotropic fields that correspond in particular to the case of cosmological fields in redshift space (where the galaxy distribution is distorted along the line-of-sight).

The beauty of these estimators resides in their ability to extract cosmological information from the topological features of different observed fields (galaxy distribution, lensing maps, CMB). In section 5.1.1.1, we will show how they can be used to extract information on dark energy and modified gravity models.

2.3 Birth and growth of the cosmic web

Understanding the formation and growth of the cosmic web has important implications in both galaxy formation and cosmology. Indeed, the large-scale structure is meant to be one of the major probes of cosmology in the coming years by means of for instance gravitational lensing, structure growth factor, baryon acoustic oscillations and void analysis. The cosmic web is also paramount to understanding the role of the environment in galaxy formation and evolution. It induces a density bias and preferred directions that should have an impact on galaxies. In particular, as will be discussed later, it has indeed been shown that halo's properties are correlated to their environment (see section 4.1.2).

2.3.1 Lagrangian perturbation theories and the Zel'dovich approximation

An interesting approach to understand the genesis of the cosmic web is to adopt a Lagrangian view dual to the Eulerian description used in section 2.1.2. Instead of expressing quantities at fixed comoving coordinates (Eulerian view), one can indeed write the cosmic fields as a function of the initial position $\mathbf{q}$. This initial position is related to the comoving coordinate $\mathbf{x}$ at time t by a displacement term

$$\mathbf{x}(t) = \mathbf{q} + \psi(\mathbf{q}, t) . \tag{2.79}$$

Starting from an homogeneous initial density field, the local density at time t then reads

$$\rho(\mathbf{q}, t) = \frac{\bar{\rho}}{J} , \tag{2.80}$$

where J is the jacobian of the Eulerian-to-Lagrangian transformation $J = \left|\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\mathbf{q}}\right|$ given by

$$J = \left|\delta_{ij}^K + \frac{\partial \psi_i}{\partial q_j}\right| , \tag{2.81}$$

with δ_{ij}^K the Kronecker delta. The Euler equation for $\mathbf{u} = a\dot{\mathbf{x}}$, then yields

$$\ddot{\psi} + 2H\dot{\psi} = -\frac{1}{a^2}\nabla_{\mathbf{x}}\Phi , \tag{2.82}$$

where $\nabla_{\mathbf{x}}\Phi$ is seen as a function of $\mathbf{q}$. Using the Poisson equation to express the gravitational potential in terms of the density, it yields

$$\nabla_{\mathbf{x}}\ddot{\psi} + 2H\nabla_{\mathbf{x}}\dot{\psi} = -4\pi G\bar{\rho}\delta = 4\pi G\bar{\rho}\frac{J-1}{J} , \tag{2.83}$$

The linear solution for the displacement field is often referred to as the Zel'dovich approximation ([Zel'dovich, 1970]). In this regime, equation 2.83 becomes

$$\nabla_{\mathbf{q}}\ddot{\psi} + 2H\nabla_{\mathbf{q}}\dot{\psi} = \frac{3}{2}\Omega H^2\nabla_{\mathbf{q}}\psi , \tag{2.84}$$

as $J \approx 1 + \mathrm{tr}(\partial\psi_i/\partial q_j) = 1 + \nabla_{\mathbf{q}}\psi$ at linear order. This equation is the same as equation 2.9 so that the solution – if we neglect the decaying mode – is given by

$$\psi_{ZA}(\mathbf{q}, t) = D_+(t)\psi^+(\mathbf{q}) \tag{2.85}$$

and

$$\rho_{ZA}(\mathbf{q}, t) = \frac{\bar{\rho}}{|(1 - D_+(t)\lambda_1)(1 - D_+(t)\lambda_2)(1 - D_+(t)\lambda_3)|} , \tag{2.86}$$

with $\lambda_1 > \lambda_2 > \lambda_3$ the eigenvalues of $-\partial\psi_i^+/\partial q_j$. Note that the deformation tensor $-\partial\psi_i^+/\partial q_j$ is proportional to the Hessian of the gravitational potential up to a time-dependent factor $2(3D_+a^2H^2\Omega)^{-1}$. The Zel'dovich displacement is therefore a rectilinear trajectory that moves particles along the direction of the initial force. It allows one to probe the mildly non-linear stages of structure formation and describe the anisotropic collapse of matter that shapes the cosmic web. Indeed, Equation 2.86 suggests that the first caustics, i.e the first shell crossings, appears when $D_+(t)\lambda_1 \to 1$ and therefore correspond to planar regions, the so-called Zel'dovich pancakes. This situation occurs when $\lambda_1 > 0$. If λ_2 is also positive, one direction of the wall will later collapse leading to the formation of a filament. If λ_3 is similarly positive, the filament will eventually contract to form a fully collapsed object (halo, cluster at the nodes of the cosmic web). The Lagrangian formalism described in this section is therefore a powerful descriptor of structure formation and allows one to understand how matter evolves on large scales escaping from the voids to the walls then flowing towards the filaments and finally streaming towards the nodes. This picture takes place at various scales and epochs, the speed of the successive collapses being directly encoded in the magnitude of the eigenvalues. It is quite striking that this very simple model allows one to understand the formation of the cosmic web that is seen both in observations and in simulations

It has to be noted that the Zel'dovich ballistic approximation breaks down as soon as shell crossing occurs. Indeed when two flows of matter crosses, the gravity in this high-density regions tend to capture the streams of matter and form localized structures whereas in the Zel'dovich approach, the two streams would simply pass through each other and continue their free streaming motion. Different ideas have been put forward to improve the description of the cosmic web. The adhesion model ([Gurbatov et al., 1989, Kofman et al., 1990, Shandarin, 1991, Hidding et al., 2012]) for instance proposes to add an artificial viscosity that sticks the particles in the multi-flow regions and indeed reproduces with better accuracy the cosmic web seen in numerical simulations.

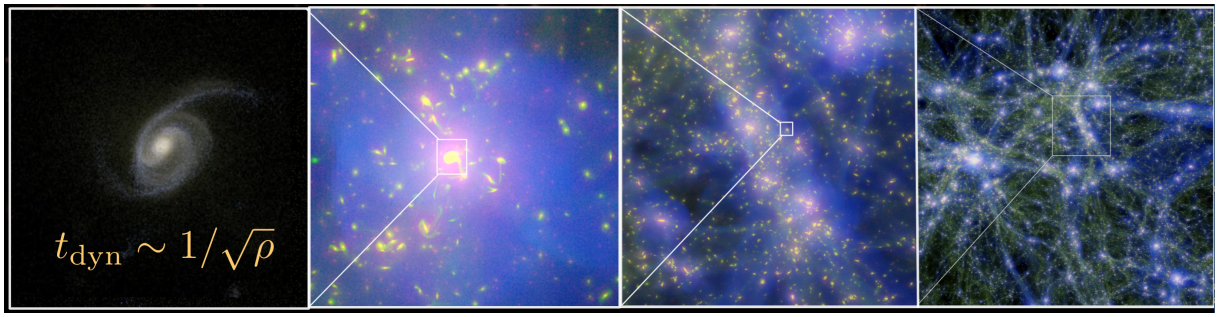


Figure 2.12: Zoom at different scales in the MareNostrum simulation (2.12) illustrating the clustering of galaxies in filaments.

2.3.2 Voids, walls, filaments and nodes

The ellipsoidal collapse model developed and studied in particular by [Lynden-Bell, 1964, Lin et al., 1965, Icke, 1973, White and Silk, 1979, Peebles, 1980, Watanabe and Inagaki, 1991, Lemson, 1993, Bond and Myers, 1996a], following Zel'dovich's ideas in the seventies allow one to understand the anisotropic nature of the gravitational collapse. In particular, they predict that elongated and flattened structures form. Those ideas led to the concept of cosmic web ([Klypin and Shandarin, 1993, Bond et al., 1996]) that was precisely described in the context of the peak-patch formalism ([Bond and Myers, 1996a, Bond and Myers, 1996b, Bond and Myers, 1996c]). The origin of filaments and nodes lies in the asymmetries of the initial Gaussian random field describing the primordial universe and amplified by gravitational collapse. The above-mentioned works pointed out the importance of non-local tidal effects in weaving the cosmic web. The high-density peaks define the nodes of the evolving cosmic web and completely determine the filamentary pattern in between. In particular, one can appreciate the crucial role played by the study of constrained random fields in understanding the geometry of the large-scale matter distribution. This will be one of the major tool used in the rest of the thesis.

The global picture that emerges is therefore the following. The clustering of matter on cosmological scales follows a complex network made of giant bubble-like voids surrounded by sheets which intersects into elongated filaments with clusters of galaxies at their nodes. This structure is seen both in numerical simulations as seen on figure 2.12 and observations (see section 3.2) and occurs at various scales. The following section will be devoted to the theoretical study of this cosmic web and in particular its various mathematical definitions.

2.3.3 The skeleton of the large-scale structure

The characterization of the cosmic web is a challenging task that first requires one to elaborate a proper mathematical definition of the complex cosmic network made of large voids, sheet-like structures and elongated filaments that our eye easily detects. In the literature, lots of different definitions and algorithms have emerged and have been applied to simulations and observations in order to extract the cosmic filamentary pattern together with its corresponding set of walls and voids.

2.3.3.1 On the plurality of morphological estimators

Some earlier attempts were based on percolation and graph theory like the minimum spanning tree ([Barrow et al., 1985]).

More recently, a class of cosmic web classifiers appeared initially designed for simulated data and relying on the use of a *smooth* field. The simplest approaches propose to identify the different components of the cosmic web from the density Hessian matrix ([Aragón-Calvo et al., 2007b]), tidal field ([Hahn et al., 2007b],[Forero-Romero et al., 2009]) or velocity shear ([Hoffman et al., 2012]) depending on the sign of their eigenvalues. In this context, the curvature of the density field is a natural indicator of the local morphology of the matter distribution and one can simply count the number of positive eigenvalues of the density Hessian matrix in order to define voids (3 positive eigenvalues), walls (2), filaments (1) and nodes (0). Recent developments by [Aragón-Calvo et al., 2007a] and [Cautun et al., 2013] extract the cosmic web from the joint analysis of the density field smoothed at different scales. The T and V-web methods classify structures from a dynamical or kinematic point of view i.e using respectively the tidal or the velocity shear fields as suggested by the Zel'dovich picture described in section 2.3.1. When the cosmic web is defined as the locus of shell crossings, a phase space analysis can be carried out in order to detect the caustics in the density field ([Neyrinck, 2012]) but requires the knowledge of the full phase space information and is therefore far from being applicable to real data.

Another criterion based on Morse theory was put forward by [Novikov et al., 2006, Sousbie et al., 2008, Pogosyan et al., 2009b] to define the local skeleton as the set of critical lines joining the maxima of the (density) field through saddle points following the gradient. It will be developed in more details in the following section. This definition therefore uses the geometry of the density field in a more sophisticated way than simple eigenvalues counting. This is also the case of [Bond et al., 2010] who uses the Hessian eigen-directions to define filaments.

However, in order to get a fully connected cosmic web, one is led to consider non-local definitions. An interesting approach was followed by [Platen et al., 2007, Platen et al., 2008, Sousbie et al., 2009, Aragón-Calvo et al., 2010] who used water-shedding to identify voids, filaments and walls. The global skeleton algorithm by [Sousbie et al., 2009] will be described in the next section.

The intrinsic point-like nature of real data have led to a second class of web identifiers initially motivated by the observations. For instance, the Candy or Bisous model ([Stoica et al., 2005]) is based on a marked point process and was recently applied to the Sloan Digital Sky Survey ([Tempel et al., 2014]). Another method proposed by [González and Padilla, 2010] relies on the positions and masses of dark matter haloes. More recently, the state-of-the-art persistent skeleton by [Sousbie, 2011, Sousbie et al., 2011] allowed the use of the skeleton formalism on point-like distributions as the result of recent developments in computational topology and in particular regarding persistent homology and discrete topology.

As a conclusion, the cosmic web can be defined in several very different ways. None of them is better than the others in all circumstances but each method is optimised for a given purpose, simulations versus observations for instance.

2.3.3.2 The skeleton picture

In the rest of this section, the skeleton algorithm is described with a particular emphasize on the local theory and the global algorithm that will be used in the works presented in part II. The global definition allows an accurate extraction of the filamentary pattern in numerical simulations. It will be used to study the impact of the large-scale environment on dark matter halos and galaxies' properties. On the other hand, the local skeleton allows one to study the theoretical properties of the cosmic web from first principles, as shown in section 2.3.3.3.

First, let us define the critical points of a field ρ as the points where $\nabla\rho = 0$. Minima (resp. maxima) are then the critical points for which λ_i, the eigenvalues of the Hessian matrix H of the field, are positive (resp. negative). The third type of critical points are the saddle points for which at least one eigenvalue is negative and one is positive (in three dimensions, there are two different types of saddle points depending on the number (1 or 2) of positive eigenvalues). If one tries to define filaments, one may draw the lines starting at the maxima of the field and following special lines which seem denser than the others, like the crest lines of a mountain. This is the basic idea used to define the skeleton.

In mathematical terms, the skeleton of a field ρ is the subset of critical lines (i.e. field lines which go through critical points) connecting the saddle points and the local maxima of ρ and parallel to its gradient. Being parallel to the gradient is equivalent to say that you are on a field line, it means that if you drop a viscous ball, this ball follows one and only one special line, a field line, defined by the direction of the gradient. This definition is actually equivalent to the following one. Let me define the peak (resp. void) patches of the density field as the set of points converging to a specific local maximum (resp. minimum) while following the field lines in the direction (resp. opposite direction) of the gradient. The skeleton is then the intersection of the void patches. So as to make this definition more intuitive and concrete, let me use the example of a landscape, as we often do in the context of Morse theory. In this 2D example, maxima are the peaks of the mountains, minima are the lake floors. To understand the definition of a peak patch, let us pour a tin of very high visibility green paint from a peak, the paint will color a part of the mountain and the border of this patch will pass through minima and saddle points. This patch is the set of points reached by a field line converging to the given maxima, it means a peak patch. On the contrary, let us color in red all the points of the landscape from where, when one pours a drop of paint, the paint flows until reaching a given minimum, this is the void patch attached to this minimum. The skeleton is then the borders of the void patches, in other words the crest lines: being on a crest line, a step on one side or on the opposite one, makes one fall in one valley or the another i.e. towards one minimum or the other. If this definition seems very natural, it is also strongly non-local. However, it is possible to build a local approximation of the global skeleton.

2.3.3.3 The local skeleton

By Taylor expanding the field, a second-order approximation of the skeleton is obtained and called the local skeleton ([Novikov et al., 2006, Sousbie et al., 2008, Pogosyan et al., 2009b]) that we propose to recap here. The local skeleton is the set of points where the modulus of the gradient of the density field is minimum along an isocontour. In 3D, if one wants to describe a closed surface (like an isodensity contour) with a coordinate system (u, v), it is not possible to find one which is not singular in some points. That is why, three coordinate systems s_i must be defined so that, if the isocontour is $(x_1(u, v); x_2(u, v); x_3(u, v))$, then while s_i is varying, one

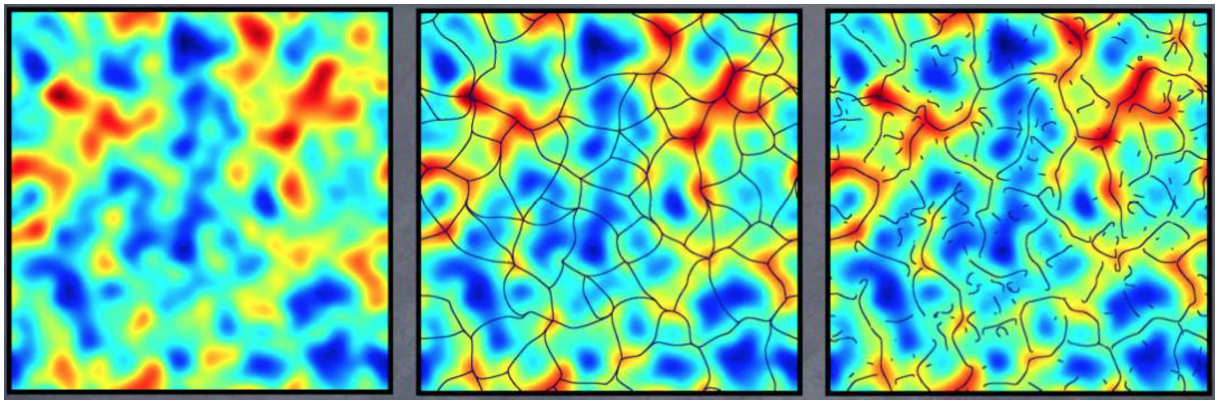


Figure 2.13: *Left panel*: a 2D density map where red color stands for high densities and blue color for low-densities. *Middle panel*: the 2D global skeleton (black solid line) links maxima through saddle points like the crest line of a landscape. *Right panel*: the 2D local skeleton (black solid line) is a local – but not connected – approximation of the global one.

stays in the plane $(x_j; x_k)$ for $i \neq j \neq k$. The condition of isodensity and of extremal gradient can then be written to obtain for $i \neq j$

$$\begin{bmatrix} S_i \\ S_j \end{bmatrix} = 0 \tag{2.87}$$

with

$$S_i = \frac{\partial^2\rho}{\partial r_j \partial r_k}\left(\frac{\partial\rho^2}{\partial r_j} - \frac{\partial\rho^2}{\partial r_k}\right) + \frac{\partial\rho}{\partial r_j}\frac{\partial\rho}{\partial r_k}\left(\frac{\partial^2\rho}{\partial r_k^2} - \frac{\partial^2\rho}{\partial r_j^2}\right) - \frac{\partial\rho}{\partial r_i}\left(\frac{\partial\rho}{\partial r_k}\frac{\partial^2\rho}{\partial r_i \partial r_j} - \frac{\partial\rho}{\partial r_j}\frac{\partial^2\rho}{\partial r_i \partial r_k}\right) . \tag{2.88}$$

The skeleton condition given by equation 2.87 can alternatively be rewritten in terms of the Hessian matrix of the density field

$$H \cdot \nabla\rho \times \nabla\rho = 0. \tag{2.89}$$

This condition is a condition of extremal gradient, not only minimal. The local skeleton of the density field is therefore obtained via the following condition on the eigenvalues of the Hessian matrix ($\lambda_1 \geq \lambda_2 \geq \lambda_3$)

$$\lambda_2 < 0,\ \lambda_3 < 0,\ H \cdot \nabla\rho = \lambda_1 \nabla\rho\,. \tag{2.90}$$

This equation is the starting point of the numerical implementation. First, the discrete point-like distribution must be smoothed using a cloud-in-cell interpolation (sampling on a grid and sharing the mass of a particle in a cell to each vertex depending on the distance between the vertex and the particle) and convolved with a Gaussian window (to smooth on a given scale). Then, the skeleton of this density field is traced by solving equation 2.90: it can be done by searching the null isocontour of S_i with the marching cube algorithm which computes the value of S_i for the eight vertices of each cell and with this information, finds the intersection of the isocontour and the cell (a triangle). Then the solution is at the intersection of the three surfaces defined by the algorithm except when there is a singularity. Therefore, the algorithm chooses for each cell the two surfaces for which we are far from the situation $\nabla\rho$ parallel to r_i , i.e. where it is singular. In practice, it means that the largest $\det(\nabla\rho, r_i, r_j)$ is selected and the corresponding intersection computed. Then the eigenvalues of the Hessian matrix are calculated

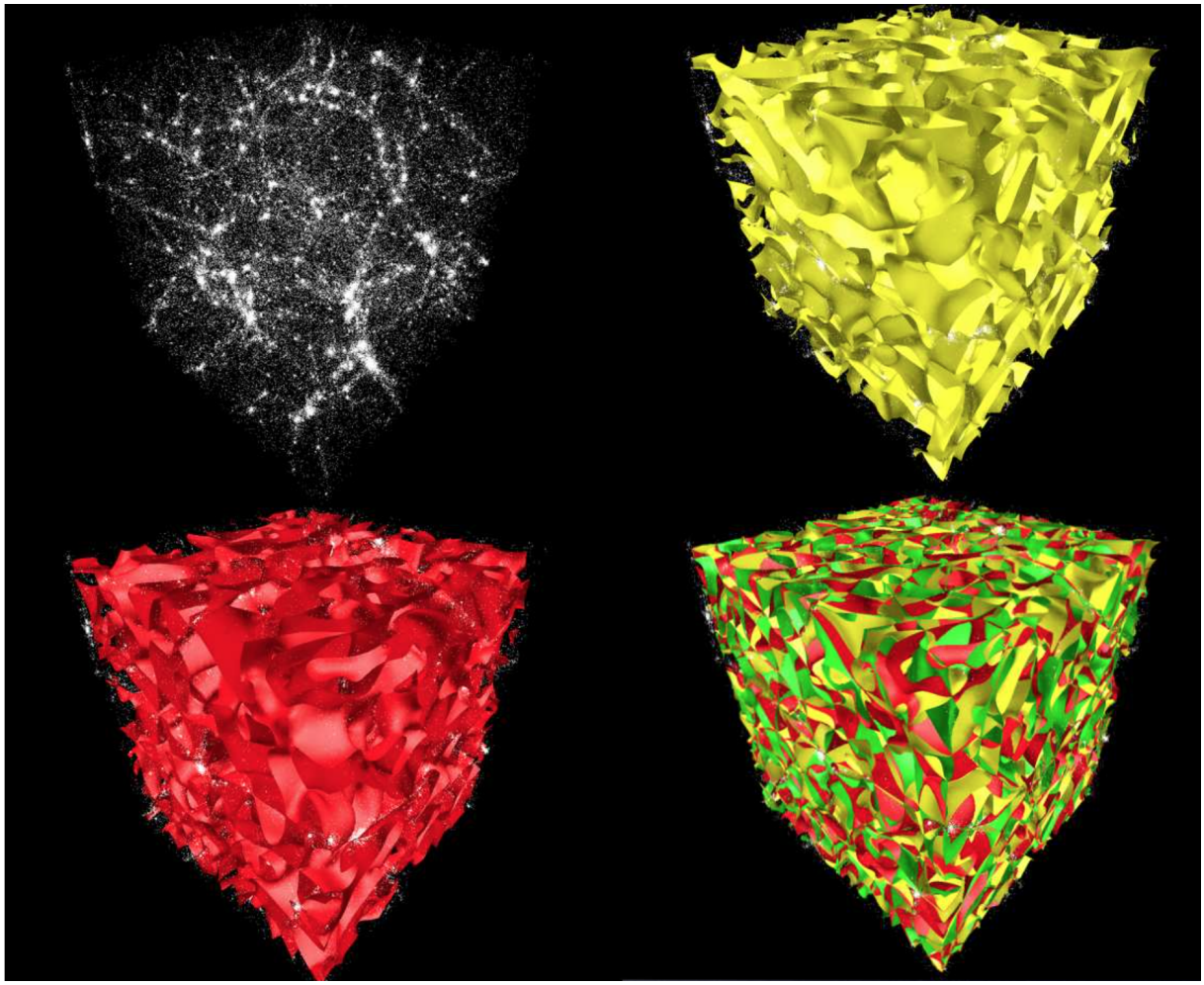


Figure 2.14: Construction of the 3D local skeleton at the intersection of three surfaces defined by $S_i = 0$.

in order to keep only the solutions where $\lambda_2, \lambda_3 < 0$. An illustration of the construction of the local skeleton is given in figure 2.14 . This local skeleton is not connected (see the right panel of figure 2.13). It is therefore best to find a solution to compute the global one. Nevertheless, it is far from being useless to study the local skeleton as the local theory of the skeleton ([Pogosyan et al., 2009b]) allows one to analytically predict the properties of the cosmic web (length of the filaments, connectivity, etc).

2.3.3.4 The global skeleton algorithm

The key-point to compute the global skeleton ([Sousbie et al., 2009]) is to solve the following equation

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \nabla\rho\,. \tag{2.91}$$

[Sousbie et al., 2009] implemented the solution of equation 2.91 by a probabilist approach to identify void patches and then filaments at the intersection of the void patches. The main idea is the following: starting at the minima of the field, all pixels are considered in increasing density

order and are given a probability of belonging to each void patches depending on their relative height to their labelled neighbours. Once finished, each pixel is assigned to its void patch of higher probability. The main advantage of this method is to provide a fully connected set of critical lines as shown in figure 2.13 but three shortcomings remain:

- Morse theory formalism deals with Morse functions, it means smooth and non-degenerate functions. As we work with point-like fields, we need to smooth the density field before computing its skeleton and therefore to introduce a smoothing length.
- this method does not account optimally for Poisson noise; indeed, a better algorithm should carry out adaptative smoothing in order to get a more precise skeleton where lots of data are available and a smoother version where data are less numerous.
- more dramatically, all watershed algorithms implemented on discretized meshes will over produce filaments as the segmentation is carried at finite resolution, where the underlying Morse theory is not satisfied.

A recent developments by [Sousbie, 2011, Sousbie et al., 2011] resolved these two issues by dealing with discrete Morse theory and introducing the notion of persistence that allows one to mimic an adaptative smoothing depending on the local level of noise.

3 Observing galaxies

In this chapter, we present some of the basics of the current paradigm of galaxy formation with a particular emphasis on spin acquisition and we describe two major observational effects: redshift space distortion and intrinsic alignments.

3.1 Galaxy formation

While the gravitational evolution of the dark matter density field is well known, observations give access to the galaxy population and there is therefore a real need to link galaxies to the underlying density field. The common picture is as follows: galaxies reside in large concentrations of dark matter called dark matter halos which typically host one central galaxy and several satellites. Those virialised halos are the highly non-linear result of gravitational dynamics. They tend to form in the high-density peaks of the density field by gravitational instability as described previously. Galaxies can therefore be associated to the peaks of the density field. Those peaks are biased tracers of the density field ([Kaiser, 1984, Bardeen et al., 1986]). In addition to peak bias, galaxy formation is highly non-linear with complex baryonic processes at play (feedback, etc). This additional complexity is called galaxy biasing ([Dekel and Rees, 1987]).

3.1.1 Dark matter halos

In this section, we propose to describe how dark halos are biased tracers of the underlying dark matter density field (section 3.1.1.1). We also characterize their mass distribution (section 3.1.1.2) and how they acquire angular momentum (section 3.1.1.3) as those two ingredients – mass and spin – are crucial to understanding galaxy morphology and therefore how the Hubble sequence is set up.

3.1.1.1 Bias

When the galaxy distribution is to be inferred from the well-known dark matter field, the simplest idea is to identify them

$$\delta_{\rm g} = \delta_{\rm DM}\,. \tag{3.1}$$

However, if galaxies are located at the peaks of the density field, they must necessarily represent a biased tracer. Indeed, building upon the idea of [Kaiser, 1984] that the density of galaxy clusters must be strongly clustered compare to the density of galaxies, the work of [Davis et al., 1985] led to the notion of linear bias

$$\delta_{\rm g} = b_1\delta_{\rm DM}\,, \tag{3.2}$$

where b_1 is a constant. Indeed [Kaiser, 1984] showed that the correlations between two regions lying above a threshold ν can be written

$$1 + \xi_{>\nu} = \frac{\int_\nu^\infty \mathrm{d}x \int_\nu^\infty \mathrm{d}y\, \mathcal{P}(x,y)}{\left(\int_\nu^\infty \mathrm{d}x\, \mathcal{P}(x)\right)^2}, \tag{3.3}$$

where

$$\mathcal{P}(x,y) = \frac{1}{2\pi\sqrt{1-\xi^2}} \exp\left[-\frac{1}{2}\frac{x^2 - 2\xi xy + y^2}{1-\xi^2}\right], \tag{3.4}$$

$\xi = \xi(r)$ being the correlation function of the density field. At large separation (i.e small ξ), equation 3.3 becomes

$$\begin{aligned} 1 + \xi_{>\nu} &= \frac{\int_\nu^\infty \mathrm{d}x\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) f\left(\frac{\nu - \xi x}{\sqrt{1-\xi^2}}\right)}{f^2(\nu)} \\ &= \frac{\int_\nu^\infty \mathrm{d}x\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \left(f(\nu) - \xi x f'(\nu) + \mathcal{O}(\xi^2)\right)}{f^2(\nu)} \\ &= 1 + \xi \frac{e^{-\nu^2/2}}{2\pi f^2(\nu)} + \mathcal{O}(\xi^2) \end{aligned}$$

where $f(x) = \frac{1}{2}\mathrm{Erfc}\left(\frac{x}{\sqrt{2}}\right)$. In the high ν, high separation limit, the correlation function between two regions lying above a threshold ν thus reads

$$\xi_{>\nu} \approx \nu^2 \xi, \tag{3.5}$$

so that the correlation function of high density regions decreases more slowly than the density field correlation function with an amplification factor or bias that is proportional to the threshold squared.

Given the complexity of galaxy formation, the validity of this approximation is likely to be very narrow. More sophisticated biasing schemes involve stochastic ([Dekel and Lahav, 1999]), time-evolving ([Nusser and Davis, 1994, Fry, 1996, Tegmark and Peebles, 1998]), non-linear ([Mo and White, 1996, Pen, 1998, Guo and Jing, 2009]), non-local ([Matsubara, 1999]), scale-dependent ([Lumsden et al., 1989, Mann et al., 1998]) bias. Indeed, a linear bias cannot be preserved by the non-linear gravitational evolution. This bias is also very likely to depend on the population of galaxies and to be not only a function of the density but also temperature, merging history and other galaxy properties, introducing some scatter in the galaxy-matter density relation. Note that the scale-dependence is strong on small scales but remains weak with typical deviations of less than a few percents on scales above 20 Mpc/h ([Crocce et al., 2013]).

To account for this bias, different approaches have been applied to galaxy surveys such as a constant bias per redshift bin or physically motivated bias models. Those bias models typically take the following functional form known as the generalized time-dependent bias model ([Clerkin et al., 2015])

$$b(z) = c + \frac{b_0 - c}{D^\alpha(z)}. \tag{3.6}$$

The issue of bias is crucial in astrophysics if one wants to infer cosmological information from galaxy observations. It is also an interesting quantity to measure on its own as it carries interesting information on the physics of galaxy formation. If analytical studies can hardly be used

as the sole tool to predict the galaxy bias, numerical simulations are definitely a complementary way of accurately predict and calibrate galaxy observables. Improvements of our understanding of galaxy formation and its implementation in simulations is therefore one of the major challenge of the coming decades for astrophysicists.

3.1.1.2 Press-Schechter mass function

While the full non-linear evolution of dark matter halos requires the use of numerical simulations, some of their properties are within reach from first principles. For instance, the statistics of the number and mass of halos can be predicted as a function of cosmic time via the Press & Schechter formalism ([Press and Schechter, 1974]). This approach relates the population of collapsed objects to the overdense peaks of the density field. Under spherical symmetry, initial peaks grow following the spherical collapse model presented in section 2.1.3 until a linearly extrapolated critical density $\delta_c \approx 1.686$. Therefore once peaks reach the critical density, they collapse to form a halo. Let us sketch here how to use those ideas to investigate the halo mass function.

Let us first study mass fluctuations in a given volume V_0 of mean density ρ_0 and mass $M_0 = \rho_0 V_0$ using a Fourier decomposition and let us call W the filter function normalised such that $\int W(\mathbf{x})\mathrm{d}\mathbf{x} = V$. The variance of the fluctuations is then $\sigma_M^2 = \left\langle (\delta M/M)^2 \right\rangle$ where the mass $M(\mathbf{x})$ in the volume V centered on $\mathbf{x}$ can be written

$$M(\mathbf{x}) = \int \rho(\mathbf{y}) W(\mathbf{x}-\mathbf{y})\mathrm{d}\mathbf{y} \tag{3.7}$$

$$= \int \rho_0(1+\delta(\mathbf{y}))W(\mathbf{x}-\mathbf{y})\mathrm{d}\mathbf{y}\,. \tag{3.8}$$

The variance of the mass fluctuations therefore reads

$$\sigma_M^2 \propto \int k^2 \mathrm{d}k P(k) W(k)^2\,. \tag{3.9}$$

For a filter that cuts large wavelengths ($> 1/R$) and a density fluctuation such that $P(k) = k^n$, this variance becomes

$$\sigma_M^2 = \int_0^{1/R} k^2 P(k)\mathrm{d}k \propto M^{-\frac{n+3}{3}}\,. \tag{3.10}$$

Hence, for $n > -3$, the amplitude of fluctuations decreases as perturbation scale grows, it means that small objects become non-linear first (i.e form first) and later aggregate to form bigger structures by merging. This bottom-up growth of structures is known as hierarchical clustering.

The Press-Schechter formalism then states that the fraction of mass clustered in halos of mass larger than $M = 4\pi R^3/3$ at redshift z and denoted $f(>M,t)$ corresponds to regions in the initial condition with (interpolated linear) density above the critical density given by the spherical collapse $\delta_c \approx 1.686$. If $\delta(k,t_0)$ is known, one can express $\delta(k,t)$ its linear interpolation at time $t > t_0$ such that $\delta(k,t) = D_+(t)/D_+(t_0)\delta(k,t_0)$, then $\delta M(M,t) = D_+(t)/D_+(t_0)\delta M(M,t_0)$ is the mass fluctuations at a given scale M. Assuming that initial conditions are Gaussian, it yields

$$f(>M,t) = \int_{\delta_c}^{\infty} \mathrm{d}\delta \frac{1}{\sqrt{2\pi}D(t)\sigma_M(M,t_0)} \exp\left(-\frac{\delta^2}{2D^2(t)\sigma_M(M,t_0)}\right), \tag{3.11}$$

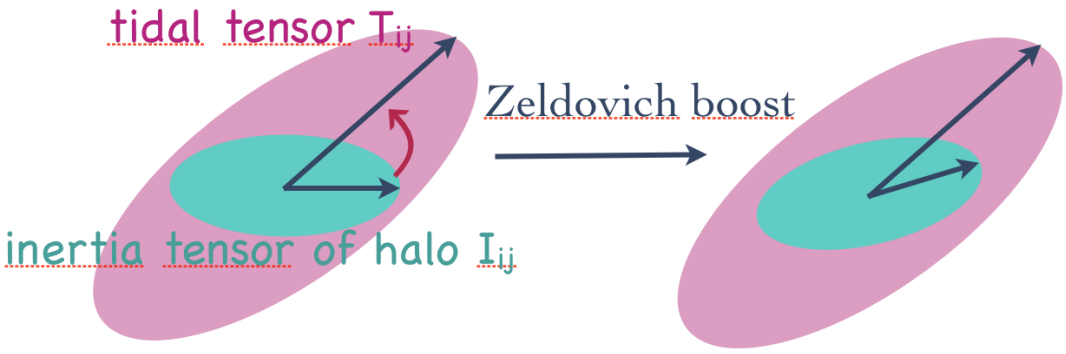


Figure 3.1: Spin acquisition by tidal torquing. At linear order, the misalignment between the inertia tensor of the proto-object and the surrounding tidal tensor induces an inhomogeneous Zel'dovich boost which corresponds to the acquisition of a net intrinsic angular momentum in Eulerian space.

where for the sake of simplicity we denote $D(t) = D_+(t)/D_+(t_0)$. This can simply be rewritten

$$f(>M,t) = \int_{\delta_c/D\sigma_M}^{\infty} \mathrm{d}u \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) = \frac{1}{2}\left[1 - \mathrm{Erf}\left(\frac{\delta_c}{\sqrt{2}D(t)\sigma_M(M,t_0)}\right)\right] . \tag{3.12}$$

According to Press and Schechter, $f(>M,t)$ represents the fraction of space which has collapsed in objects of mass at least M. However when M tends to zero i.e. σ_M tends to infinity, this fraction of space tends to $1/2$ while it should be one. Consequently, they decided to arbitrarily multiply f by 2. Assuming this extra factor of 2, the expected number density of objects of mass M, $n(M)$ can be related to f by $f(M) = \int_M^\infty n(m)m\,\mathrm{d}m/\rho_0$ or equivalently $n(M) = -f'(M)\rho_0/M$. Therefore,

$$n(M,t) = -\sqrt{\frac{2}{\pi}}\frac{\rho_0}{M}\frac{\delta_c}{\sigma^2(M,t)D(t)}\frac{\mathrm{d}\sigma(M,t)}{\mathrm{d}M}\exp\left(-\frac{\delta_c^2}{2D^2(t)\sigma^2(M,t)}\right) . \tag{3.13}$$

There is an exponential cutoff in the halo mass function for which the associated mass is given by $M_\star(t)$ such that $\delta_c = D(t)\sigma(M_\star(t),t)$. This mass is called the Press-Schechter mass and characterizes the typical mass that is collapsing at a given time t. At the current epoch, most of the mass is now inside galaxies. This mass function reproduces with good accuracy the result of numerical simulations despite this ad hoc factor of two. The Press-Schechter formalism was later extended to the excursion set formalism ([Bond et al., 1991]) and to ellipsoidal collapse ([Sheth and Tormen, 2002]).

3.1.1.3 Spin acquisition

In the standard paradigm of galaxy formation, protogalaxies acquire their spin by tidal torquing coming from the surrounding matter distribution ([Hoyle, 1949, Peebles, 1969, Doroshkevich, 1970, White, 1984, Catelan and Theuns, 1996, Crittenden et al., 2001]). At linear order, this spin is acquired gradually until the time of maximal extension (before collapse) and is proportional to the misalignment between the inertia tensor of the protogalaxy and the surrounding tidal tensor (see [Schaefer, 2009] for a review)

$$L_i = a^2(t)\dot{D}_+(t)\sum_{j,k,l}\epsilon_{ijk}I_{jl}T_{lk} , \tag{3.14}$$

where $a(t)$ is the scale factor, D_+ the growth factor, T_{ij} the tidal tensor (detraced Hessian of the gravitational potential), I_{ij} the protogalactic inertia tensor (only its traceless part, $\overline{I}_{ij}$ contributes to the spin). Indeed, the spin of a protogalaxy contained in a volume V and with center of gravity located at position $\bar{\mathbf{r}}$ can be written

$$\mathbf{L} = \int_V \mathrm{d}^3\mathbf{r}(\mathbf{r}-\bar{\mathbf{r}}) \times (v(\mathbf{r}) - v(\bar{\mathbf{r}}))\rho(\mathbf{r})\,, \tag{3.15}$$

where the implicitly time-dependent velocity field is denoted $v(\mathbf{r})$ and the mass density $\rho(\mathbf{r})$. In Lagrangian coordinates, equation 3.15 becomes

$$\mathbf{L} = \rho_0 a^5 \int_{V_L} \mathrm{d}^3\mathbf{q}(\mathbf{x}-\bar{\mathbf{x}}) \times (\dot{\mathbf{x}}(\mathbf{q}) - \dot{\mathbf{x}}(\bar{\mathbf{q}}))\,. \tag{3.16}$$

In the Zel'dovich approximation where $\mathbf{x} = \mathbf{q} - D_+\nabla\psi(\mathbf{q})$ and $\dot{\mathbf{x}} = -\dot{D}_+\nabla\psi$, ψ being the displacement field (such that $\Delta\psi = \delta$), and assuming that the gradient of the displacement field is almost constant across the proto-object of Lagrangian volume V_L, equation 3.16 can be evaluated by a second-order Taylor expansion

$$\mathbf{L} \approx -\dot{D}_+\rho_0 a^5 \int_{V_L} \mathrm{d}^3\mathbf{q}(\mathbf{q}-\bar{\mathbf{q}}) \times [T\cdot(\mathbf{q}-\bar{\mathbf{q}})]\,, \tag{3.17}$$

where T is the tidal shear tensor $T_{ij} = \partial_i\partial_j\Psi_{ij}$ at the center of gravity. Let us define the inertia tensor I

$$I_{ij} = \rho_0 a^3 \int_{V_L} \mathrm{d}^3\mathbf{q}(q_i - \bar{q}_i)(q_j - \bar{q}_j)\,, \tag{3.18}$$

so that the spin of the proto-galaxy can eventually be written as equation 3.14. It is mainly advected until the time of maximal extension before collapse as the lever arm is then drastically reduced. This process of spin acquisition by tidal torquing is illustrated on figure 3.1. It is clear from equation 3.14 that spherical proto-objects can not acquire angular momentum in this context but it can be shown ([Peebles, 1969]) that spin acquisition in spherically symmetric settings is possible as a second-order effect.

In the Lagrangian picture, I_{ij} is the moment of inertia of a uniform mass distribution within the Lagrangian image of the halo, while T_{ij} is the tidal tensor averaged within the same image. Thus, to rigorously determine the spin of a halo, one must know the area from which matter is assembled, beyond the spherical approximation. While this can be determined in numerical experiments, theoretically we do not have the knowledge of the exact boundary of a protohalo. As such, one inevitably has to introduce an approximate proxy for the moment of inertia (and an approximation for how the tidal field is averaged over that region).

The most natural approach is to consider that protohaloes form around an elliptical peak in the initial density and approximate its Lagrangian boundary with the elliptical surface where the over-density drops to zero. This leads to the following approximation for the inertia tensor ([Catelan and Theuns, 1996, Schäfer and Merkel, 2012])

$$I_{ij} = \frac{M}{5}\begin{pmatrix} A_y^2 + A_z^2 & 0 & 0 \\ 0 & A_z^2 + A_x^2 & 0 \\ 0 & 0 & A_x^2 + A_y^2 \end{pmatrix}, \tag{3.19}$$

(in the frame of the Hessian) where the mass of the protohalo is $M = 4/3\pi A_x A_y A_z \rho_0 a_0^3$ and the semi-axes of the ellipsoid, A_i, are function of the eigenvalues of the Hessian (negative for a

peak),

$$A_i = \sqrt{\frac{2\nu\sigma_2}{-\lambda_i}}\,. \tag{3.20}$$

with ν the peak overdensity. Note that the traceless part of I_{ij} that is relevant for torques is then proportional to the traceless part of the inverse Hessian of the density field

$$\overline{I}_{ij} = \frac{2}{5}\nu\sigma_2 M \overline{H}_{ij}^{-1}\,. \tag{3.21}$$

This scenario of spin acquisition by linear tidal torquing has been tested against numerical simulations. It appears that the spin (in particular its direction) is well-predicted in the early stages of galaxy evolution ($z < 3$) before non-linearities beyond Zel'dovich become important ([Efstathiou and Jones, 1979, Barnes and Efstathiou, 1987, Sugerman et al., 2000, Porciani et al., 2002a, Porciani et al., 2002b, Lee and Pen, 2008]). In particular, they confirm two predictions of linear tidal torque theory : the spin scales like the mass to the power 5/3 and the spin parameter $\lambda = L/\sqrt{GRM^3}$ is anti-correlated with the peak height (the higher the peak, the lower the spin parameter). However at later times, both spin direction and magnitude deviate significantly from the linear prediction.

3.1.2 Where baryons come into play

Galaxies reside inside dark matter halos. They are made of dark matter but also baryons (dust, gas, stars) and are therefore sensitive to much more complex physical processes than collisionless cold dark matter. As developed in section 2.1.4.2, on small scales (typically below a few megaparsecs), baryonic effects need to be taken into account. This requires to modeling baryonic processes like feedbacks from stars and black holes. Eventually, the complete understanding of galaxy formation is still a burning issue in astrophysics. One of the central question is to understand how the Hubble sequence is set up and in particular what are the main drivers of galaxy formation and evolution: internal processes or environmental effects (nature vs nurture)?

3.1.3 The role of the environment

It is now well-established that some of the properties of galaxies are observed to be correlated with their environment. For example, morphology is not independent from the local density: elliptical galaxies are found preferentially in dense regions and spiral galaxies are found in the field ([Oemler, 1974, Dressler, 1980, Hermit et al., 1996, Guzzo et al., 1997]). This type of correlation between galaxies and environment is similarly observed for their colour, luminosity, surface brightness or spectral type ([Norberg et al., 2002, Hogg et al., 2003, Blanton et al., 2005, Lee and Lee, 2008, Lee and Li, 2008, Tempel et al., 2011, Yan et al., 2012, Kovač et al., 2014]). Observations also indicate that the rotation axes of galaxies are correlated with the filaments in which they are embedded ([Navarro et al., 2004, Trujillo et al., 2006, Lee and Erdogdu, 2007, Paz et al., 2008, Jones et al., 2010, Tempel et al., 2013, Zhang et al., 2013, Tempel and Libeskind, 2013]).

Studying the properties of galaxies as a function of their location in the cosmic web yields valuable information about the formation and evolution of galaxies. For this reason, it is necessary to use simulations to investigate how galaxies and halos are correlated to their environment. It was found by [Hahn et al., 2007b, Hahn et al., 2007a, Hahn et al., 2009, Gay et al., 2010, Metuki

et al., 2015] that the properties of dark matter halos such as their morphology, luminosity, colour and spin parameter depend on their environment as traced by the local density, velocity and tidal field. On top of scalar quantities, it also appeared that their shape and spin were correlated to the directions of the surrounding filaments and walls both in dark matter ([Aubert et al., 2004, Bailin and Steinmetz, 2005, Brunino et al., 2007, Hahn et al., 2007b, Hahn et al., 2007a, Aragón-Calvo et al., 2007b, Paz et al., 2008, Zhang et al., 2009, Codis et al., 2012, Libeskind et al., 2013, Aragon-Calvo and Yang, 2014]) and hydrodynamical simulations ([Navarro et al., 2004, Hahn et al., 2010, Dubois et al., 2014]). The impact of the environment on galaxy formation will be developed further in the works presented in section 4.

3.2 Observational systematics

The comparison between real observations and theoretical predictions is obviously not straightforward. Different observational effects contaminate the signal : completeness, masks, instrumental errors among many others. In this section, we will focus on two major observational systematics that will be central in the rest of the thesis : redshift space distortions and intrinsic alignments of galaxies.

3.2.1 Redshift space distortion

3D surveys are conveyed in redshift space where the hypothesis of isotropy breaks down. Indeed in astrophysical observations, the three dimensional positions of structures are frequently not accessible directly. While angular positions on the sky can be obtained precisely, the radial line-of-sight (hereafter LOS) position of objects is determined by proxy, e.g., via measurements of the LOS velocity component. This implies that galaxy distribution data are presented in redshift space in cosmological studies. Figures 3.2 and 3.3 illustrates the amplitude of such redshift distortion as a function of scale. On small, non-linear scales, the well-known finger-of-God effect ([Jackson, 1972, Peebles, 1980]) stretches the collapsing clusters along the line-of-sight whereas on larger scales (see figure 3.3) the redshift space distortion flattens the voids along the LOS ([Sargent and Turner, 1977]), in accordance to the linear result of [Kaiser, 1987].

The estimation of position via redshift assigns to a given object the "redshift" coordinate, $\mathbf{s}$,

$$\mathbf{s} = \mathbf{r} + H^{-1}\mathbf{v} \cdot \hat{\mathbf{r}} \tag{3.22}$$

shifted from the true position $\mathbf{r}$ by the projection of the peculiar velocity $\mathbf{v}$ along the line-of-sight direction $\hat{\mathbf{r}}$.

On large scales, in the linear regime of density evolution, the mapping to redshift coordinates induces an anisotropic change in mass density contrast ([Kaiser, 1987]), best given in Fourier space

$$\hat{\delta}^{(s)}(\mathbf{k}) = (1 + f\mu^2)\hat{\delta}^{(r)}(\mathbf{k}), \tag{3.23}$$

that has dependency on the angle $\mu = \mathbf{k} \cdot \hat{\mathbf{r}}/k$ between the direction of the wave $\mathbf{k}$ and the line of sight, and the amplitude, f, tracing the growth history of linear inhomogeneities $D(a)$, $f = \mathrm{d}\log D/\mathrm{d}\log a \approx \Omega_m^{0.55}$, ([Peebles, 1980]). The main qualitative effect of this distortion is the enhancement of clustering via the squeezing overdense regions and the stretching underdense voids along the line of sight. If matter is traced by biased halos (e.g. galaxies), the redshift

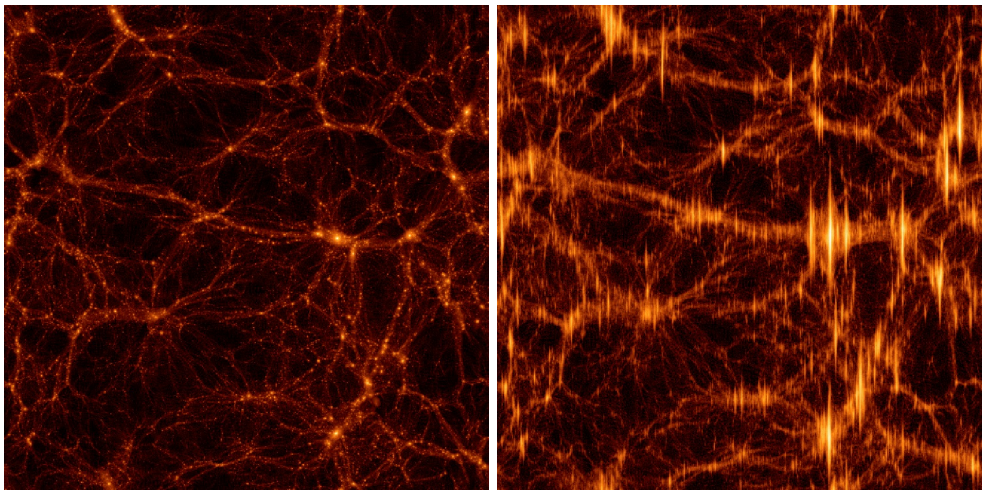


Figure 3.2: *Left panel*: an example of a slice through a 512^3 dark matter particles ΛCDM simulations at redshift 0 ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.92$) in real space. The boxsize is 100 Mpc/h, the slice thickness 10 Mpc/h. *Right panel* shows the same field when redshift distortion has been applied along the ordinate. Fingers of God are quite visible on that slice (see also figure 3.3).

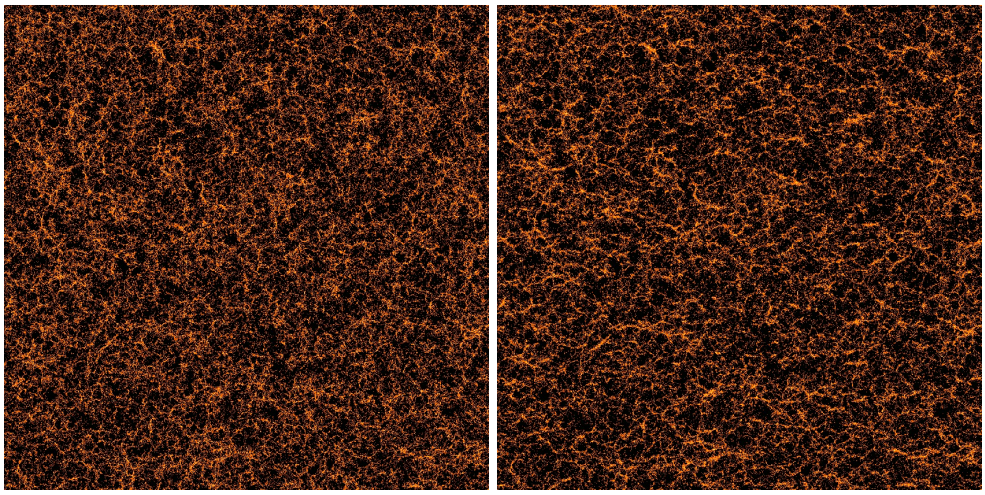


Figure 3.3: A slice through the Horizon 4π halo catalog at redshift zero without (*left panel*) and with (*right panel*) redshift distortion (along the ordinate). The boxsize is 2 Gpc/h, the slice thickness 40 Mpc/h. The area is 20×20 larger than the slice presented in figure 3.2. Each dot represents a halo colour-coded by its mass. Note the clear preferred horizontal elongation of structures in redshift space.

space distortion can be modeled using a linear bias b_1 factor ([Kaiser, 1984]) so that the galaxy contrast in redshift space reads

$$\hat{\delta}_g^{(s)}(\mathbf{k}) = (1 + \beta\mu^2)\hat{\delta}_g^{(r)}(\mathbf{k})\,, \quad \beta \equiv f/b_1\,. \tag{3.24}$$

At linear order, the power spectrum in redshift space can therefore be written in terms of the real-space density power spectrum $P_{\delta\delta}$

$$P^s_{\rm lin}(k,\mu) = P_{\delta\delta}(k)\left(1 + 2\mu^2\beta + \mu^4\beta^2\right) \tag{3.25}$$

and all the information of the corresponding correlation function is enclosed in the coefficients of the first three even Legendre polynomials L_0, L_2 and L_4

$$\xi^s(r,\mu) = \xi_0(r)L_0(\mu) + \xi_2(r)L_2(\mu) + \xi_4(r)L_4(\mu) \tag{3.26}$$

where r is the pair separation and μ the angle between the separation vector and the line-of-sight. Equation 3.25 is valid only if the velocity field is unbiased which may not be the case in the context of peak theory for instance. Then, one can rewrite the linear redshift space power spectrum in a more general setting as

$$P^s_{\rm lin}(k,\mu) = P_{gg}(k) - 2\mu^2 P_{g\theta}(k) + \mu^4 P_{\theta\theta}(k)\,, \tag{3.27}$$

where $P_{gg} = \left\langle |\hat{\delta}_g^{(r)}\hat{(\mathbf{k})}|^2 \right\rangle$, $P_{g\theta} = \left\langle \hat{\delta}_g^{(r)}(\mathbf{k})\hat{\theta}(\mathbf{k}) \right\rangle$ and $P_{\theta\theta} = \left\langle |\hat{\theta}(\mathbf{k})|^2 \right\rangle$.

In the mildly non-linear regime, redshift space distortions interplay with non-Gaussian corrections that develop with the growth of non-linearities. [Scoccimarro et al., 1999, Bernardeau et al., 2002] established the framework for a perturbative approach to this regime. The fully non-linear expression (generalizing equation 3.23) for the Fourier transform of the density in redshift space is

$$\hat{\delta}_s(\mathbf{k}) = \int \frac{\mathrm{d}^3\mathbf{x}}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{x}} e^{if k_z v_z(\mathbf{x})} \left[\delta^{(r)}(\mathbf{x}) + f\nabla_z v_z(\mathbf{x})\right]\,, \tag{3.28}$$

with v_z the peculiar velocity along the LOS, $f = \mathrm{d}\log D/\mathrm{d}\log a$, while assuming the plane-parallel approximation and that only $f\nabla_z v_z(\mathbf{x}) < 1$ terms contribute[1]. Expanding the exponential in equation 3.28, leads to, using the kernels Z_n, the following expression for the density field in redshift space

$$\hat{\delta}_s(\mathbf{k},\tau) = \sum_{n=1}^{\infty} D_1^n(\tau) \int \mathrm{d}^3\mathbf{k}_1 \cdots \int \mathrm{d}^3\mathbf{k}_n \delta_{\rm D}(\mathbf{k} - \mathbf{k}_1 - \cdots - \mathbf{k}_n) Z_n(\mathbf{k}_1, \cdots, \mathbf{k}_n)\hat{\delta}_l(\mathbf{k}_1)\cdots\hat{\delta}_l(\mathbf{k}_n), \tag{3.29}$$

where μ_i is the cosine of the angle between $\mathbf{k}_i$ and the line of sight, $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2$. The first kernels are given by (assuming a quadratic local bias model involving b_1 and b_2)

$$Z_1(\mathbf{k}) = (b_1 + f\mu^2)\,, \tag{3.30}$$

$$Z_2(\mathbf{k}_1,\mathbf{k}_2) = b_1 F_2(\mathbf{k}_1,\mathbf{k}_2) + f\mu^2 G_2(\mathbf{k}_1,\mathbf{k}_2) + \frac{f\mu k}{2}\left[\frac{\mu_1}{k_1}(b_1 + f\mu_2^2) + \frac{\mu_2}{k_2}(b_1 + f\mu_1^2)\right] + \frac{b_2}{2}, \tag{3.31}$$

[1] Note that equation 3.28 could be more accurately replaced by $\hat{\delta}_s(\mathbf{k}) = -\delta_{\rm D}(\mathbf{k}) + 1/(2\pi)^3 \int \mathrm{d}^3\mathbf{x} e^{-i\mathbf{k}\cdot\mathbf{x}} e^{if k_z v_z(\mathbf{x})} \left[\delta^{(r)}(\mathbf{x}) + 1\right]$ where no assumption about the amplitude of the radial velocity is made. This gives exactly the same perturbation theory as expected.

with $f = \Omega_m^{\gamma_m}$, $\gamma_m \simeq 6/11$ and $\epsilon \simeq 3/7\,\Omega_m^{-2/63}$. In particular

$$\hat{\delta}_s^{(1)}(\mathbf{k}) = D_1(\tau) \int \mathrm{d}^3\mathbf{k}_1 \delta_\mathrm{D}(\mathbf{k}-\mathbf{k}_1) Z_1(\mathbf{k}_1)\hat{\delta}_l(\mathbf{k}_1) = D_1(\tau) Z_1(\mathbf{k})\hat{\delta}_l(\mathbf{k})\,, \tag{3.32}$$

$$\hat{\delta}_s^{(2)}(\mathbf{k}) = D_1^2(\tau) \int \mathrm{d}^3\mathbf{k}_1 \int \mathrm{d}^3\mathbf{k}_2 \delta_\mathrm{D}(\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2) Z_2(\mathbf{k}_1,\mathbf{k}_2)\hat{\delta}_l(\mathbf{k}_1)\hat{\delta}_l(\mathbf{k}_2). \tag{3.33}$$

The difference compared to the linear regime resides in the coefficients $A_{0,1,2,3}$ entering the expansion $P^s(k,\mu) = \sum_{i=0}^3 A_i \mu^{2i}$. Unfortunately, because redshift space distortions introduce important couplings between small and large scales, a naive perturbation theory in redshift space has very bad convergence properties and can only describe scales exceeding several tens of megaparsecs.

It is in particular necessary to model the large-scale perturbative regime and the small-scale fingers of God at the same time leading to the study of streaming models ([Scoccimarro, 2004]) that take into account the distribution of random pair velocities in collapsed object through a multiplicative factor $D(k,\mu^2)$

$$P^s_\mathrm{NL}(k,\mu) = D(k,\mu^2) P^s_\mathrm{lin}(k,\mu)\,. \tag{3.34}$$

Gaussian and Lorentz dampings have been mainly considered in the literature to describe the small-scale velocity dispersion

$$D_\mathrm{Gaussian}(k,\mu^2) = \exp\left(-(k\sigma\mu)^2\right)\,, \tag{3.35}$$

$$D_\mathrm{Lorentz}(k,\mu^2) = \left(1+(k\sigma\mu)^2\right)^{-1}\,, \tag{3.36}$$

where σ is the rms velocity dispersion. Improvements of this model include for instance a recent work by [Taruya et al., 2010] allowing almost percent precision at BAO scale which is necessary to achieve percent precision on the determination of the cosmological parameters.

It has to be noted that redshift space distortions are not only a source of nuisance but also can contain interesting information. Indeed, measuring redshift space distortions in galaxy surveys is believed to be a promising way to study the large-scale structure and constrain dark energy and modified gravity models through the growth rate of structure formation and the large-scale velocity field. Correlation functions and power spectra have been extensively used in this context ([Peacock et al., 2001, Percival et al., 2004, Guzzo et al., 2008, Blake et al., 2011a]) and should play a significant role in future large surveys.

3.2.2 Intrinsic alignments

Weak lensing is often presented as a potential powerful probe of cosmology for the coming years with large surveys like DES, Euclid or LSST. It relies on the fact that the picture of galaxies that we observe in the sky is distorted because the light path from background sources towards us is bent by the gravitational potential well along the line of sight. The apparent deformation of the shapes of galaxies on the sky can therefore be used to map the gravitational potential or measure its statistical properties like its power spectrum so that it eventually allows to probe cosmology (cosmological model, dark matter distribution, etc). The idea is thus to try and detect coherent distortions of the shapes of galaxies e.g using the two-point correlation function of the ellipticities of galaxies.

3.2.2.1 II and GI contaminations

The projected surface mass density, integrated along the line of sight to distant sources is often preferred over the potential although they are trivially related by the Poisson equation. The effective convergence κ, which is nothing but the dimensionless projected density, is statistically described by its power-spectrum $P_\kappa(\ell)$ as a function of wavenumber ℓ. For a source at comoving distance χ_s, we can write the convergence at an angular position θ ([Bartelmann and Schneider, 2001])[2]:

$$\begin{aligned}\kappa(\theta,\chi_s) &= \frac{1}{c^2}\int_0^{\chi_s} \mathrm{d}\chi\, \frac{(\chi_s-\chi)\chi}{\chi_s}\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right]\Phi, \\ &= \frac{3H_0^2\Omega_0}{2c^2}\int_0^{\chi_s}\mathrm{d}\chi\,\frac{(\chi_s-\chi)\chi}{\chi_s}\frac{\delta(\chi\theta,\chi)}{a(\chi)},\end{aligned} \tag{3.37}$$

where χ is the comoving distance, a the expansion factor, δ, the density contrast, and Φ, the three-dimensional gravitational potential, are related by the Poisson equation

$$\Delta\Phi = \frac{3H_0^2\Omega_0}{2a}\delta\,. \tag{3.38}$$

This can easily be generalised to a population of sources with a broad redshift distribution ([Bartelmann and Schneider, 2001]).

The relation between Φ and δ can be cast into a relation between the lensing potential ϕ and effective convergence κ, and the effective shear γ_i, which involves the traceless parts of the projected tidal tensor. All these quantities are defined by:

$$\phi = \frac{2}{c^2}\int_0^{\chi_s}\mathrm{d}\chi\,\frac{(\chi_s-\chi)\chi}{\chi_s}\Phi\,,\quad \kappa = \frac{1}{2}\left(\frac{\partial^2\phi}{\partial\theta_1^2}+\frac{\partial^2\phi}{\partial\theta_2^2}\right), \tag{3.39}$$

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\phi}{\partial\theta_1^2}-\frac{\partial^2\phi}{\partial\theta_2^2}\right)\,,\quad \gamma_2 = \frac{\partial^2\phi}{\partial\theta_1\partial\theta_2}\,. \tag{3.40}$$

It is generally suitable to treat the shear in complex notations $\gamma=\gamma_1+i\gamma_2$. This quantity is most easily accessible as it captures the amount of anisotropic distortion a light bundle experiences on its way from a distant source to the observer. Therefore, the observed ellipticity of such a source, in the weak lensing regime of small distortions, is directly related to the shear. Indeed, by also defining a complex ellipticity $e=e_1+ie_2=|e|e^{2i\psi}$, such that $|e|=(1-q)/(1+q)$ and $q=b/a$ is the major (a) to minor (b) axis ratio, we have

$$e = e_s+\gamma\,, \tag{3.41}$$

where e is the apparent ellipticity and e_s the intrinsic source ellipticity (the one we would have observed without lensing).

An important statistics of this cosmic shear distortion field is the two-point correlation of projected ellipticities that can formally be split into the following components:

$$\langle e(\vartheta)e(\vartheta+\theta)\rangle_\vartheta = \langle e_s e_s'\rangle + 2\langle e_s\gamma'\rangle + \langle\gamma\gamma'\rangle\,, \tag{3.42}$$

where, for compactness, the prime means at an angular distance θ from the first location. The cosmological weak lensing signal is commonly decomposed into the ξ_+ and ξ_- shear correlation functions. Following [Schneider et al., 2002], $\xi_\pm$ is given by

$$\xi_\pm(\theta) = \langle\gamma_+\gamma_+\rangle \pm \langle\gamma_\times\gamma_\times\rangle = \frac{1}{2\pi}\int_0^\infty \mathrm{d}\ell\,\ell P_\kappa(\ell)J_{0/4}(\ell\theta)\,,$$

[2] Simplified to a flat Universe case

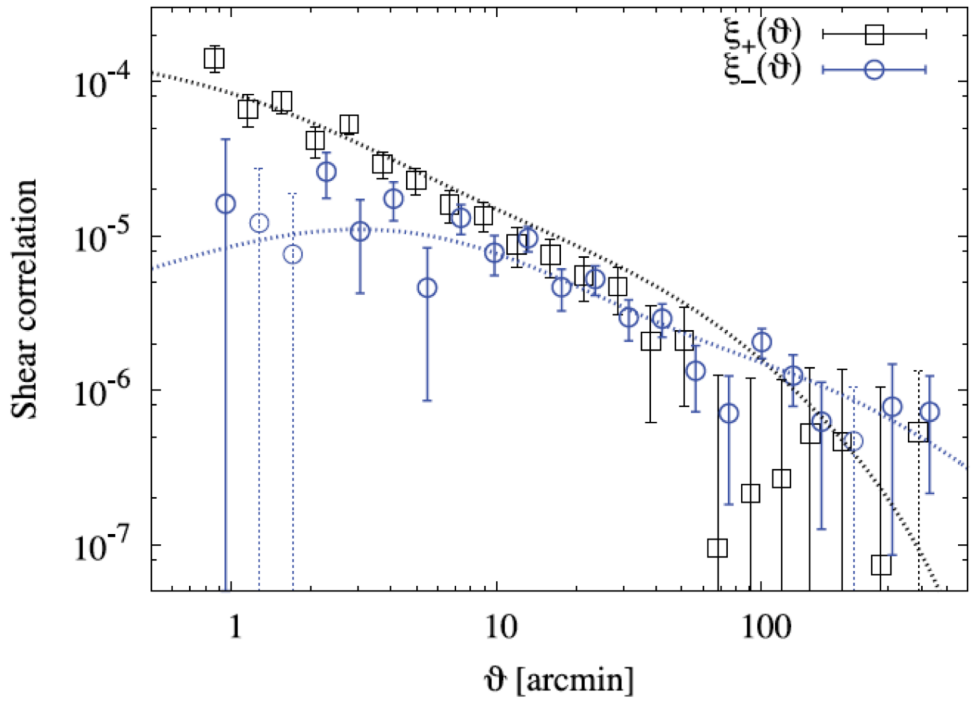


Figure 3.4: ξ_+ and ξ_- shear correlation functions as measured in the CFHTLenS data by [Kilbinger et al., 2013].

where J_0 and J_4 are the 0-th and 4-th order Bessel functions for ξ_+ and ξ_- respectively. In this expression, γ_+ (resp. $\gamma_\times$) is the component of the complex shear orientated $0/90^o$ (resp. $\pm 45^o$) with respect to the line connecting two galaxies separated by a projected distance θ. Figure 3.4 displays the ξ_+ and ξ_- shear correlation functions as measured by [Kilbinger et al., 2013] in the CFHTLenS data.

The fundamental assumption of weak lensing, which allows one to infer shear properties from observed ellipticities is that, on average, the intrinsic orientation of sources is completely random. The breakdown of this hypothesis yields additional terms to $\langle \gamma\gamma' \rangle$ on the right-hand side of equation (4.1) that have to be carefully accounted for in observations. The weak lensing signal is therefore contaminated by the two kinds of intrinsic alignments (IA):

- the so-called "II" term $\langle e_s e'_s \rangle$ induced by the intrinsic correlation of the shape of galaxies in the source plane ([Heavens et al., 2000, Croft and Metzler, 2000, Catelan et al., 2001]). This mostly concerns pairs of galaxies that are at similar redshifts.

- and the so-called "GI" term $\langle e_s \gamma' \rangle$ coming from correlation between the intrinsic ellipticity of a galaxy and the induced ellipticity (or shear) of a source at higher redshift ([Hirata and Seljak, 2004, Heymans et al., 2006, Joachimi et al., 2011]). This non-trivial term is indirectly explained if the shape of galaxies is correlated with the local gravitational tidal field, which also contributes to the shear signal experienced by the far source in a given pair of observed ellipticities.

3.2.2.2 Observing intrinsic alignments

Much effort has thus been made to control the level of IA of galaxies as a potential source of systematic errors in weak gravitational lensing measurements although some techniques have been proposed to mitigate their nuisance by making extensive use of photometric redshifts (e.g. [Bridle and King, 2007, Joachimi and Schneider, 2008, Joachimi and Schneider, 2010, Joachimi and Bridle, 2010, Kirk et al., 2010, Blazek et al., 2012]). Direct measurements of the alignment of the projected light distribution of galaxies in wide field imaging data seem to agree on a contamination at a level of a few percents in the shear correlation functions, although the amplitude of the effect depends on the depth of observations, the amount of redshift information and the population of galaxies considered in the sense that red galaxies seem to show a strong intrinsic projected shape alignment signal whereas observations only place upper limits in the amplitude of the signal for blue galaxies (e.g. [Brown et al., 2002, Lee and Pen, 2002, Bernstein and Norberg, 2002, Heymans et al., 2004, Hirata et al., 2004, Mandelbaum et al., 2006, Hirata et al., 2007, Mandelbaum et al., 2011, Joachimi et al., 2011, Joachimi et al., 2013a]).

From the theoretical point of view, it has been shown that dark halos ([Aragón-Calvo et al., 2007b, Paz et al., 2008, Codis et al., 2012] among many others) and galaxies ([Hahn et al., 2010, Dubois et al., 2014]) are correlated with the cosmic web (see also section 4.1.2). The consequence of this large-scale coherence of galaxies could then contaminate significantly the weak lensing observables. Numerical works based on halo model and semi-analytical modelling (e.g. [Heavens et al., 2000, Croft and Metzler, 2000, Schneider and Bridle, 2010, Schneider et al., 2012, Joachimi et al., 2013b]) or hydrodynamical simulations ([Codis et al., 2015a, Tenneti et al., 2015]) together with analytical models (e.g. [Catelan et al., 2001, Hirata and Seljak, 2004]) have therefore tried to examine this effect. Here we give a brief overview of the field but refer the interested reader to the extensive reviews of [Joachimi et al., 2015, Troxel and Ishak, 2015, Kiessling et al., 2015] for further details.

3.2.2.3 Analytical modelling

Analytical models are all based on the assumption that intrinsic alignments are tidally generated leading to the so-called tidal alignment models that are typically twofold : a linear model for elliptical galaxies that can be seen as a tidally distorted stellar spheroid and a quadratic model for spirals whose spin is believed to emerge from tidal torquing. It has to be noted that linear here means linear in the tidal field. Other sources of non-linearities include the non-linear dynamics of the dark matter field and non-linear galaxy bias. The linear model have for instance been amended so as to take into account the non-linear evolution of the dark matter field, leading to the so-called non-linear alignment model. This is the basis of the tidal alignment models.

As described in section 3.1.1.3, in the early stage of galaxy formation, the gradient of gravitational forces across a protogalaxy leads to angular momentum generation and tidal stretching of the shape of the object, both effects being proportional to the misalignment between the inertia tensor and the tidal tensor.

Linear model In the linear model ([Catelan et al., 2001, Hirata and Seljak, 2004]), the intrinsic ellipticity of a (typically elliptical) galaxy is proportional to the tidal field, i.e second derivatives of the gravitational potential smoothed on some scale Ψ, with a coefficient C_1 that measures the

strength of the alignment occurring at redshift $z_{\rm IA}$

$$\begin{pmatrix} e_s^+ \\ e_s^\times \end{pmatrix} = -\frac{C_1}{4\pi G} \begin{pmatrix} \nabla_x^2 - \nabla_y^2 \\ 2\nabla_x \nabla_y \end{pmatrix} \Psi(z_{\rm IA}) \,, \tag{3.43}$$

as it is assumed that stars and dark matter are in dynamical equilibrium and therefore the stellar distribution follows the distortion of the galaxy halo spheroid which is tidally distorted. The strength coefficient was measured in the SuperCOSMOS field $C_1 \approx 5 \cdot 10^{-14} (h^2\ \mathrm{M}_\odot \mathrm{Mpc})^{-1}$ ([Bridle and King, 2007]). This ansatz can be used to predict the GI and II signals for a linearly evolved potential field. If this model seems to be reasonable on scales above 10 Mpc/h ([Blazek et al., 2011]), it can be significantly improved on smaller scales by taking into account the non-linear dark matter evolution ([Bridle and King, 2007]) and other non-linear contributions like non-linear bias and the effect of weighting by the local density of galaxies ([Blazek et al., 2015]).

Quadratic model For spiral galaxies that have an angular momentum supported disc, one probably needs to use the prediction for the spin in terms of the tidal fields given by the tidal torque theory (equation 3.14). Parametrizing the correlation between inertia and shear tensor by a factor C, [Lee and Pen, 2000] were able to give a prescription for the two-point correlation function of the spin which was then improved to give a prescription for the two-point correlation function of the normalised spin vector denoted s ([Lee and Pen, 2001])

$$\langle s_\alpha s_{\alpha'} \rangle = \frac{1 + a_{\rm T}}{3} \delta_{\alpha\alpha'} - a_{\rm T} \sum_\sigma \hat{T}_{\alpha\sigma} \hat{T}_{\sigma\alpha'} \tag{3.44}$$

where $\hat{T}$ is the normalised traceless tidal shear tensor and $a_{\rm T}$ is a parameter between 0 and 3/5. Assuming that the ellipticity of the spiral is given by the projection of a circular disc perpendicular to the spin, one can compute the intrinsic ellipticity correlation function and show that ([Crittenden et al., 2001])

$$\xi_+^{II} \propto a_{\rm T}^2 \xi_{\delta\delta}^2. \tag{3.45}$$

3.2.2.4 Intrinsic alignments in simulations

While understanding intrinsic alignments from first principles can give interesting highlights on large scales, it seems however necessary to combine these analytical models with numerical simulations in order to be able to predict intrinsic alignments in the strong non-linear regime.

Early investigations focused on dark matter only simulations assuming that galaxy shapes were following dark halo shapes ([Heavens et al., 2000]). Then halo model and semi-analytical modelling have been used to assign galaxies to their host halo ([Schneider and Bridle, 2010, Joachimi et al., 2013b]). It is only recently that hydrodynamical simulations have been able to describe a sufficient large volume with enough resolution so as to measure how numerical galaxies were intrinsically aligned. Recent works by [Codis et al., 2015a, Tenneti et al., 2015] have paved the way for this investigation but it has to be remembered that hydrodynamical simulations nevertheless suffer from a lack of knowledge about galaxy formation meaning that lots of work need to be done in order to understand how much the predicted intrinsic alignments depend on sub-grid recipes and numerical schemes.

Part II

Results and Publications

4 Intrinsic alignments: numerical and theoretical insights

In this chapter, I will show how both dark matter and hydrodynamical simulations predict that the morphology of galaxies is correlated with the cosmic web. This large-scale coherence of galaxy shapes could possibly induce some non-negligible level of contamination for future cosmic shear experiments. I will show how this contamination can be measured in simulations. On the theoretical side, I will describe an analytical Lagrangian model that reproduces qualitatively the correlations between the intrinsic angular momentum of galaxies and the cosmic filaments. The key ingredient is to take into account the anisotropy of the cosmic web in the standard theory of spin acquisition by tidal torquing.

Publications

4.1 Overview

4.1.1 Introduction

Weak lensing is often presented as a potentially powerful probe of cosmology for the coming years with large surveys like DES [1], Euclid ([Laureijs et al., 2011]) or LSST [2]. It relies on the fact that the picture of galaxies that we observe in the sky is distorted because the light path from background sources towards us is bent by the gravitational potential well along the line of sight. Therefore measuring these distortions directly probes cosmology (cosmological model, dark matter distribution, etc). The idea is thus to try and detect coherent distortions of the shapes of galaxies e.g using the two-point correlation function of the ellipticities of galaxies. At this stage, it is important to realize that the apparent ellipticity of a galaxy is induced by the cosmic shear γ (which is related to the projected gravitational potential along the line of sight) but also encompasses the intrinsic ellipticity of that galaxy $e = e_s + \gamma$, where e is the apparent ellipticity and e_s the intrinsic source ellipticity (the one we would have observed without lensing). Therefore the (projected) ellipticity-ellipticity two-point correlation function can be written as the sum of a shear-shear term, intrinsic-intrinsic and intrinsic-shear correlations (see also section 3.2.2)

$$\langle e(\vartheta)e(\vartheta+\theta)\rangle_\vartheta = \langle \gamma\gamma' \rangle + \langle e_s e_s' \rangle + 2\langle e_s\gamma' \rangle \ , \tag{4.1}$$

where, for compactness, the prime means at an angular distance θ from the first location. These last two contributions that somehow contaminate the shear signal are the two kinds of intrinsic alignments, one term being the so-called "II" term $\langle e_s e_s' \rangle$ induced by the intrinsic correlation of the shape of galaxies in the source plane ([Heavens et al., 2000, Croft and Metzler, 2000, Catelan et al., 2001]) and the other one is the so-called "GI" term $\langle e_s\gamma' \rangle$ coming from correlations between the intrinsic ellipticity of a galaxy and the induced ellipticity (or shear) of a source at higher redshift ([Hirata and Seljak, 2004]).

From a theoretical point of view, there is clear evidence that the shape of galaxies is correlated to the cosmic web. This large-scale coherence should therefore induce some level of intrinsic alignments and contaminate weak lensing. In this chapter, I propose to describe an anisotropic extension of the tidal torque theory that allows to take into account the effect of the large-scale filamentary structure on galactic spin acquisition. I will also show that dark matter and hydrodynamical simulations predict that galactic spins are indeed correlated with the cosmic web in a redshift and mass-dependent way consistent with the above-mentioned anisotropic theory of tidal torques. These correlations lead to some level of intrinsic alignments that will be explicitly measured in the hydrodynamical Horizon-AGN simulation.

4.1.2 Spin acquisition within the cosmic web

Over the last decade, it has been shown that dark halos ([Aragón-Calvo et al., 2007b, Paz et al., 2008, Codis et al., 2012] among many others) and galaxies ([Hahn et al., 2010, Dubois et al., 2014]) are correlated with the cosmic web. In [Codis et al., 2012] (see paper 4.2), I indeed found in numerical simulations that the spin of dark halos is correlated with the direction of the closest filament (see also the left-hand panel of figure 4.1). Those results were obtained on the 43 million dark matter halos of the Horizon 4π simulation and are shown to be mass-dependent. Low-mass halos tend to have a spin aligned with the filament axis while the more massive ones tend

[1] http://www.darkenergysurvey.org
[2] `http://www.lsst.org`

to have a spin perpendicular to the filaments. The transition occurs around $M_{\rm tr} \approx 5 \cdot 10^{12} M_{\odot}$, is redshift-dependent, scaling as $M_{\rm tr}(z) \approx M_0(1+z)^{-\gamma_s}$ with $\gamma_s = 2.5 \pm 0.2$, and varies weakly with the scale defining filaments.

This signal can be qualitatively understood in the context of hierarchical structure formation, where small galaxies form first and merge together to form larger ones. Indeed, the first generation of small galaxies form when the walls wind up to give rise to the filaments and during this process protogalaxies acquire spin aligned with the filaments. When they later merge inside filaments by catching up with each other during their race towards the nodes, they naturally form bigger objects and convert their orbital angular momentum into spin. This spin will tend to be perpendicular to the filament axis as the impact parameter is aligned itself with the direction of the filament as sketched in the bottom right-hand panel of figure 4.2. In both cases, the key ingredient is that the acquired spin is perpendicular to the motion direction, which is in the plane of the wall and perpendicular to the forming filament in the first case and aligned with the filament in the second one.

The emergence of a vorticity field aligned with the filaments and responsible for the spin orientation of low-mass objects has been studied in [Laigle et al., 2015]. In that work (see paper A.1), we found that vorticity is confined in filaments, as expected since vorticity can only be generated in multi-flow regions and is indeed statistically aligned with the filaments. The typical cross-section perpendicular to a filament is quadrupolar, meaning that four quadrants of opposite vorticity orientation are found. The frontiers of those four quadrants coincide nicely with the walls' direction (with an excess probability of alignment of about 15%), in full agreement with the winding-of-walls scenario (see top and bottom left-hand panels of figure 4.2): matter escapes from the walls towards the forming filament before flowing inside the filament and reaching the nodes. This dynamics creates a vorticity flow aligned with the filament which is partitioned into four regions with different vorticity direction in order to respect a global zero vorticity. A filament is typically the crossing of two walls that delimitate the frontiers of the above-mentioned four regions.

More recently, we investigated in [Dubois et al., 2014] (see paper A.2) how this scenario transpires to virtual galaxies. For that purpose, we used the Horizon-AGN run, a full-physics cosmological simulation based on the hydrodynamical code Ramses and found that galactic spins are also correlated with the direction of the filaments. Low-mass, blue, disk-like, star forming galaxies tend to have a spin aligned with the filaments while massive, red, elliptical, quiescent galaxies are more likely to have a spin perpendicular. This result is illustrated on the right-hand panel of figure 4.1. It has to be noted that recent observations (in particular on the Sloan Digital Sky Survey) have reported some hints of a correlation between galaxy morphology and the cosmic web as predicted by dark matter and hydrodynamical simulations (e.g. [Tempel and Libeskind, 2013]).

4.1.3 An anisotropic theory of spin acquisition

The correlations between spins and cosmic web that are predicted by numerical simulations and confirmed by recent observations can be understood in the context of tidal torque theory [3] when the dynamical influence of filaments on galactic scales is accounted for. The key idea developed in paper 4.3 ([Codis et al., 2015b]) is to make use of the fact that walls and filaments

[3] As described in section 3.1.1.3, protogalaxies are believed to acquire most of their spin in the linear stage of structure formation –before collapse– due to the tidal torques induced by the surrounding matter distribution.

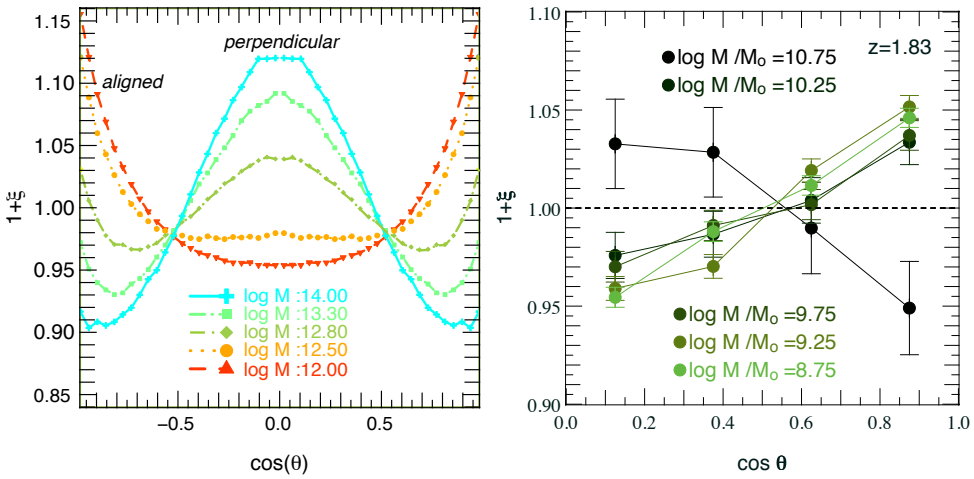


Figure 4.1: Left-hand panel: excess probability of alignment between the spin and the direction of the closest filament as measured from the 43 millions halos of the Horizon 4π simulation ([Teyssier, 2002]) at redshift zero. Different colours correspond to different mass bins from 10^{12}(red) to 10^{14} $M_\odot$ (blue) as labelled. A transition mass is detected at $M_{\rm crit}(z=0) \simeq 5(\pm1)\times10^{12}M_\odot$: for halos with $M > M_{\rm crit}$, the spin is more likely to be perpendicular to their host filament, whereas for halos with $M < M_{\rm crit}$, the spin tends to be aligned with the closest filament. This figure is from [Codis et al., 2012] (see section 4.2). Right-hand panel: same as left panel for the 160 000 galaxies of the Horizon-AGN hydrodynamical simulation at $z = 1.8$ ([Dubois et al., 2014]).

are the interference patterns of primordial density fluctuations on large scales, and induce a corresponding anisotropic boost in overdensity. On top of these modes, constructive interferences of high frequency modes produce peaks which thus get a boost in density that allows them to pass the critical threshold necessary to decouple from the overall expansion of the Universe, as envisioned in the spherical collapse model (see section 2.1.3). This well-known biased clustering effect has been invoked to justify the clustering of galaxies around the nodes of the cosmic web ([White et al., 1988]). It also explains why galaxies form in filaments: in walls alone, the actual density boost is typically not sufficiently large to trigger galaxy formation. The main nodes of the cosmic web are where galaxies migrate, not where they form. They thus inherit the anisotropy of their birth place as spin orientation. During migration, they may collide with other galaxies/halos and erase part of their birth heritage when converting orbital momentum into spin via merger. In this context, I proposed to re-visit tidal torque theory by explicitly taking into account the anisotropy of this filamentary environment on various scales in order to model primordial and secondary spin acquisition. This anisotropic tidal torque theory (ATTT) is completely analytic and relies on the study of constrained primordial random fields, namely the gravitational potential and its successive derivatives (see paper 4.3).

The angular momentum generated by linear tidal torquing is traditionally written as (see section 3.1.1.3)

$$L_i \propto \epsilon_{ijk} I_{jl} T_{lk}\,, \tag{4.2}$$

where T_{ij} is the tidal tensor (detraced Hessian of the gravitational potential) and I_{ij} the protogalactic inertia tensor (only its traceless part, $\overline{I}_{ij}$ contributes to the spin). In order to build a local theory, one is led to adopt some local proxies for the inertia tensor. In this work, I assumed that the Hessian of the density field is an accurate approximation as protogalaxies form at the peaks of the density field and should therefore get a shape that is highly correlated with the

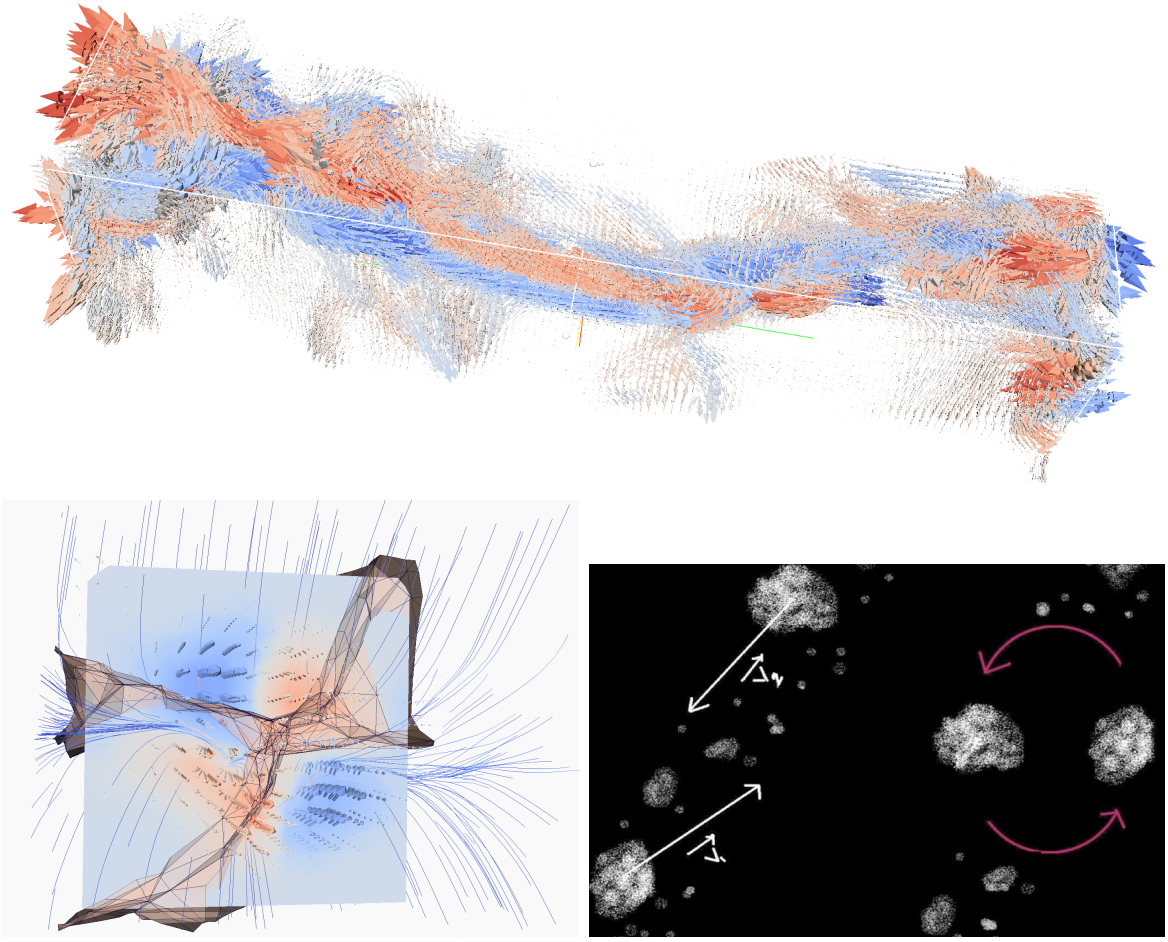


Figure 4.2: Top panel: Vorticity field along a filament colour-coded depending on its algebraic component parallel to the filament. Red arrows are going towards the left-hand side of the filament with blue arrows are pointing towards the righ-hand side of that filament. Bottom left-hand panel: Typical cross section perpendicular to the filament. Bottom right-hand panel: Sketch of the merger of two halos along the filament. The orbital angular momentum is converted into an additional spin for the resulting object, that is perpendicular to the motion direction i.e perpendicular to the filament axis.

shape of the density peak i.e with the second derivatives of the density field at the peak location. I therefore chose to study the following spin proxy

$$s_i \equiv \epsilon_{ijk} H_{jl} T_{lk} \,. \tag{4.3}$$

This approximation allows to have a local prescription for the spin that only depends on the second (by means of the tidal tensor) and fourth derivatives (Hessian of the density field) of the gravitational potential. Note that we assume here that all the fields are rescaled by their variance. In the initial conditions where the fields are supposingly Gaussian, the joint PDF of the tidal tensor and the Hessian of the density field (that we gather in a vector $\mathbf{X} = (\{T_{ij}\}_{i\leq j}, \{H_{ij}\}_{i\leq j})$) is known and simply given by

$$\mathcal{P}(\mathbf{X}) = \frac{\exp\left(-\frac{1}{2}\mathbf{X}^{\mathrm{T}} \cdot \mathbf{C_0}^{-1} \cdot \mathbf{X}\right)}{\det|\mathbf{C_0}|^{1/2}\,(2\pi)^{(\mathrm{d}+1)(\mathrm{d}+2)/2}} \,, \tag{4.4}$$

where $\mathbf{C} \equiv \langle \mathbf{X} \cdot \mathbf{X}^{\mathrm{T}} \rangle$ is the covariance matrix that can be easily computed. For instance, in two dimensions, if $\mathbf{X} = (T_{11}, T_{12}, T_{22}, H_{11}, H_{12}, H_{22})$

$$\mathbf{C} = \begin{pmatrix} 3/8 & 0 & 1/8 & -3\gamma/8 & 0 & -\gamma/8 \\ 0 & 1/8 & 0 & 0 & -\gamma/8 & 0 \\ 1/8 & 0 & 3/8 & -\gamma/8 & 0 & -3\gamma/8 \\ -3\gamma/8 & 0 & -\gamma/8 & 3/8 & 0 & 1/8 \\ 0 & -\gamma/8 & 0 & 0 & 1/8 & 0 \\ -\gamma/8 & 0 & -3\gamma/8 & 1/8 & 0 & 3/8 \end{pmatrix} ,$$

with $\gamma = \frac{\sigma_1^2}{\sigma_0 \sigma_2}$ $(= \sqrt{(n+2)/(n+4)}$ for a scale-invariant density power-spectrum $P(k)$ with spectral index n) and variances are given by

$$\sigma_n^2 = \int_0^\infty \int_0^{2\pi} \mathrm{d}\theta \; \mathrm{d}k \; P(k) k^{2n+1} \,.$$

In this context, it is straightforward to see that the mean spin is $\langle s_i \rangle = \epsilon_{ijk} \langle H_{jl} \rangle \langle T_{lk} \rangle = 0$ (all the terms entering this antisymmetric contraction are products of H_{jl} and T_{lk} that are uncorrelated e.g $H_{12}T_{23}$), as expected since there is no preferred direction on average.

However, galaxies are not forming everywhere but preferentially in filaments and nodes which define the so-called cosmic web. The origin of these structures lies in the asymmetries of the initial Gaussian random field describing the primordial universe, amplified by gravitational collapse. The presence of such large-scale structure (walls, filaments, nodes) induces local preferred directions for both the tidal tensor and the inertia tensor of forming objects which will eventually turn into preferred alignments of the spin w.r.t the cosmic web. It is therefore of interest to understand what is the expected spin direction (defined by equation 4.3) given the presence of a typical filament nearby. As a filament is the field line that joins two maxima of the density field through a filament-type saddle point (where the gradient is null and the density Hessian has two negative eigenvalues), it is of interest to study the expected spin direction of proto-objects *in the vicinity of a filament-type saddle point* with a given geometry (which imposes the direction of the filament and the wall). This constrained spin $\langle s_i(\mathbf{r}) | \mathrm{saddle} \rangle$ depends on $\mathbf{r}$, the separation vector between the proto-object and the saddle point (or equivalently the peak in 2D if we consider cross sections perpendicular to the filament). This quantity can be computed from the

joint PDF of the second to fourth derivatives of the potential field in two different locations separated by $\mathbf{r}$. Note that here we need to introduce the rescaled gradient of the density field (third derivatives of the potential) $\mathbf{q} = \nabla\delta/\sigma_1$ in order to impose a critical point constraint $\delta_D(x-\nu)|\det H_{ij}|\delta_D(q_i)$. The geometry of the critical point can also be constrained to have given values of the density curvatures i.e given eigenvalues for the Hessian λ_i. Let us gather the tidal tensor, gradient and Hessian of the density field in a vector $\mathbf{X}$ at the proto-object location (where the spin is computed) and $\mathbf{Y}$ at the critical point location (that mimics the presence of a large-scale filament). This two-point JPDF reads

$$\mathcal{P}(\mathbf{X},\mathbf{Y}) = \frac{1}{\sqrt{\det|2\pi\mathbf{C}|}}\exp\left(-\frac{1}{2}\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}^{\mathrm{T}}\cdot\mathbf{C}^{-1}\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}\right), \tag{4.5}$$

where $\mathbf{C}_0 \equiv \langle \mathbf{X}\cdot\mathbf{X}^{\mathrm{T}}\rangle$, $\mathbf{C}_\gamma \equiv \langle \mathbf{X}\cdot\mathbf{Y}^{\mathrm{T}}\rangle$ and

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_0 & \mathbf{C}_\gamma\\ \mathbf{C}_\gamma^{\mathrm{T}} & \mathbf{C}_0\end{bmatrix}.$$

Again, those covariance matrices can be computed for any power spectrum $P(k)$. For instance in two dimensions, the coefficients of those matrices read

$$\left\langle \partial_1^{i_1}\partial_2^{i_2}\phi, \partial_1^{j_1}\partial_2^{j_2}\phi\right\rangle = \int_0^\infty\int_0^{2\pi} \mathrm{d}\theta\ \mathrm{d}k\ P_k(k)\exp(\imath k\, r\cos\theta)$$
$$\imath^{i_1+i_2}(-\imath)^{j_1+j_2}(\cos\theta)^{i_1+j_1}(\sin\theta)^{i_2+j_2}\frac{k^{i_1+i_2+j_1+j_2+1}}{\sigma_{i_1+i_2}\sigma_{j_1+j_2}}. \tag{4.6}$$

Eventually, the expected spin direction of proto-objects in the vicinity of a filament-type saddle point with a given geometry $\langle s_i(\mathbf{r})|\mathrm{saddle}\rangle$ can be analytically computed from the knowledge of the joint PDF. The result – which is completely analytical – is given in paper 4.3. In short, in the plane of the saddle point, spins are aligned with the filament's direction and form four quadrants of opposite direction. When moving towards the nodes, the spins become more and more perpendicular to the filament. This geometry is illustrated in figure 4.3.

The anisotropic extension of tidal torque theory (ATTT) presented here therefore predicts a quadrupolar geometry inside filaments in agreement with the typical four quadrants of opposite vorticity seen in simulations ([Laigle et al., 2015]). It is striking that such a Lagrangian framework captures naturally the arguably non-linear Eulerian process of spin flip via mergers. In [Laigle et al., 2015], we showed that angular momentum generation of halos could be captured in Eulerian space via the secondary advection of vorticity which the formation of the filament generates, whereas in the context of tidal torque theory, we show that it may also be described in Lagrangian space via the analysis of the anisotropic tides generated by the filament to be. No description is more fundamental than the other but are the two (Eulerian versus Lagrangian) sides of the same coin. The mapping between the two descriptions requires a reversible time integrator, such as the Zeldovich approximation, which clearly limits its temporal validity to weakly non-linear scales.

Our proxy for the spin is an approximation which seems to quantitatively capture the relevant physics. It is remarkable that such an (admittedly approximate) straightforward extension of tidal torque theory captures what seems to be the driving process of spin orientation acquisition and its initial evolution.

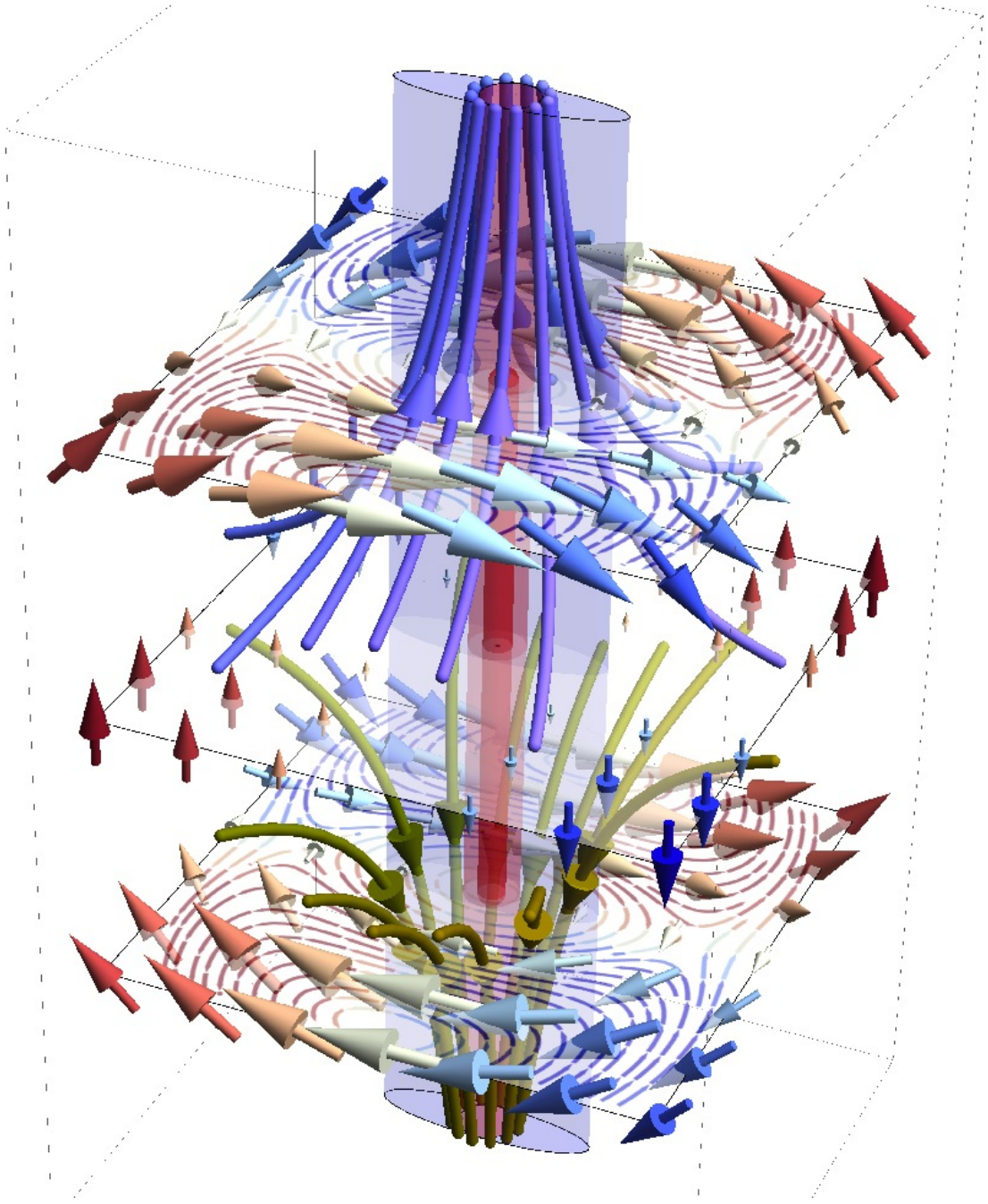


Figure 4.3: The velocity and spin flow near a vertical filament (in red) embedded in a (purple) wall (extracted from [Codis et al., 2015b]) with a flattened saddle point in the center. The purple and green flow lines trace the (Lagrangian) 3D velocities (upwards and downwards respectively). The red and blue arrows show the spin 3D distribution, while the three horizontal cross sections show spin flow lines in the corresponding plane. Note that the spin is along $\mathbf{e}_z$ in the plane of the saddle point and along $\mathbf{e}_\phi$ away from it, and that it rotates in opposite direction above and below the plane of the saddle point.

Indeed, on top of the four quadrants of opposite spin direction, it also predict a *spatial* transition for the spin direction: aligned with the filament's axis close to the saddle point and perpendicular towards the nodes of the cosmic web. This transition can be transposed into a transition in mass as seen in numerical simulations (low-mass spins aligned with the filaments and high-mass spins perpendicular) as soon as one realises that the local mass distribution of halos is expected to vary along the large-scale filament due to changes in the underlying long-wave density. In the linear regime, the typical density near the end points (nodes) of the filament, where it joins the protocluster regions, may exceed the typical density near the saddle point by a factor of two ([Pogosyan et al., 1998]). During epochs before the whole filamentary structure has collapsed, this leads to a shift in the hierarchy of the forming halos towards larger masses near the filament end points (the clusters) relative to the filament middle point (the saddle). This can be easily understood using the formalism of barrier crossing ([Peacock and Heavens, 1990, Bond et al., 1991, Paranjape et al., 2012, Musso and Sheth, 2012]), which associates the density of objects of a given mass to the statistics for the random walk of halo density as the field is smoothed with decreasing filter sizes. This formalism allows to attribute a typical mass for the objects forming in regions of a given large-scale density boost so that in the end, ATTT predicts a mass-dependent orientation of the spin in agreement with what has been measured in simulations. The transition mass, $\mathcal{M}_{\rm tr}$ for spin flip ($\langle \cos\theta \rangle = 0.5$) is found to be of the order of $4 \cdot 10^{12} M_{\odot}$, assuming a smoothing scale of 5 Mpc/h, as used in [Codis et al., 2012]. This mass is in qualitative agreement with the transition mass found in that paper, all the more so as the redshift evolution of this transition mass will also be consistent (scaling as the mass of non-linearity).

4.1.4 Intrinsic alignments of virtual galaxies

The above-described large-scale coherence of galaxies could contaminate significantly the weak lensing observables. Given the inherently anisotropic nature of the large-scale structure and its complex imprint on the shapes and spins of galaxies together with the dependency on the physical properties of the galaxies seen in the observation, it is probably difficult to rely on isotropic linear theory (e.g. [Lee and Pen, 2001]) or dark matter-only numerical simulations as the sole resort to predict and control intrinsic alignments for weak lensing applications. With the advent of cosmological hydrodynamical simulations, we are now in a position to try and measure intrinsic alignments directly into those simulations instead of relying on halo model or semi-analytical models ([Schneider and Bridle, 2010, Joachimi et al., 2013b]). In paper 4.4 ([Codis et al., 2015a]), I used the Horizon-AGN simulation presented in [Dubois et al., 2014] at redshift $z = 1.2$ to measure the level of intrinsic alignments taking the spin as a proxy for the shape of galaxies[4].

In this context, I measured on the one hand the correlation as a function of the separation between the spin of galaxies and the surrounding tidal tensor defined as the tensor of second derivatives of the gravitational potential. At zero separation, the results obtained in [Codis et al., 2012, Dubois et al., 2014] are recovered namely that galaxies have a spin correlated with the direction of the filament (which is very well aligned with the minor direction of the tidal tensor) in a mass and colour-dependent way. Beyond one-point statistics, it is also of interest in the context of weak lensing studies to quantify how this signal persists when the tidal field at a distance r from the galaxy is considered. Because the tidal field in the vicinity of a galaxy contributes also to the lensing signal carried by more distant galaxies, it is clear that the spin – tidal tensor cross-correlation is closely related to the GI term. As expected, the spin and the tidal eigen-directions then de-correlate with increasing separation. However, whereas the signal

[4] indeed, the spin is believed to be more robust w.r.t resolution effects than inertia tensors

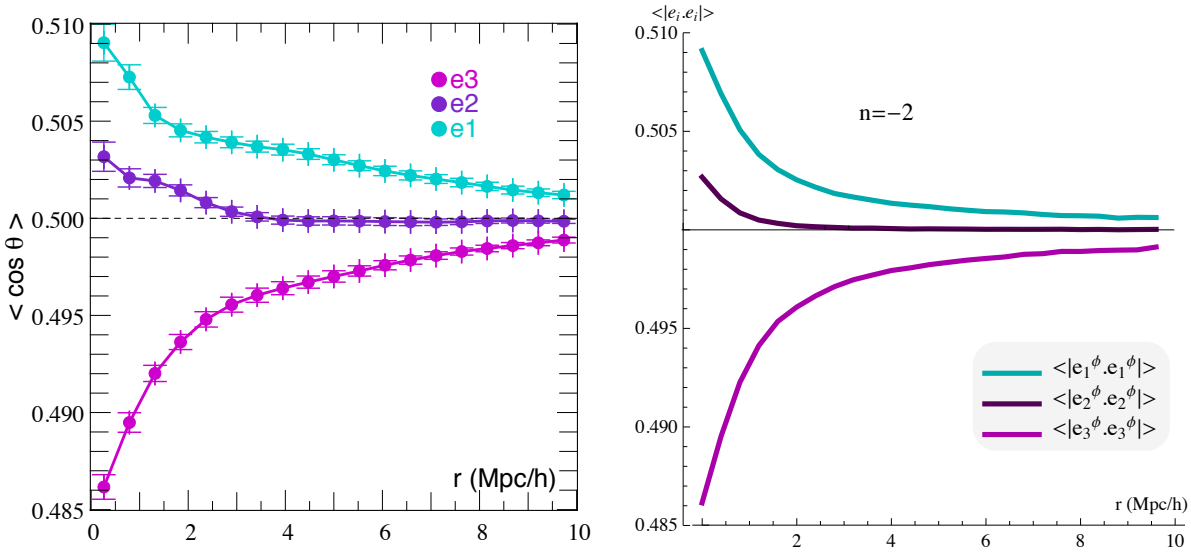


Figure 4.4: Left: Mean cosine of the angle between the spin of galaxies and the minor (cyan), intermediate (purple) and major (magenta) eigen-direction of the tidal tensor as a function of the separation in the Horizon-AGN simulation. Right: Same as left panel for a Gaussian random field with power-law power spectrum once rescaled so as to match the measured value at zero separation.

vanishes on scales $r > 3\ h^{-1}$ Mpc for the spin to intermediate tidal eigen-direction correlation, it persists on distances as large as $\sim 10\ h^{-1}$ Mpc for the minor and major eigen-directions of the tidal tensor (which is the typical size of a filament). This is potentially a worry for weak lensing experiments that could be contaminated by a quite large level of "GI" intrinsic alignments. A deeper understanding of this contamination requires being able to model projected galaxy shapes in the simulation and being as close as possible to the real observables by taking into account observational systematics like dust attenuation on galaxy colours. It is interesting to note that the behaviour of the two-point correlation function between spin and tidal tensor can be theoretically understood using a Gaussian random field δ for which we compute the joint PDF of the second derivatives of its corresponding potential (ϕ_{ij}, ϕ being related to δ by the Poisson equation). Then the mean angle between the eigen-directions of ϕ_{ij} in two locations separated by r can be computed. Once rescaled so as to match the one-point statistics (here we want to study the evolution of the two-point function with the separation, not its absolute value), we find the same qualitative behaviour as what is measured in the simulation (see figure 4.4).

On the other hand, I also measured the auto-correlations of the galaxy intrinsic ellipticities by means of the spin-spin correlation function. I found a significant spin correlation for blue galaxies out to at least 10 Mpc/h. Conversely, no significant correlations for red galaxies at that redshift ($z = 1.2$) is detected. In order to get closer to weak lensing observables, one question arises: what fraction of the spin-spin correlations remains after projection on the sky? To address this issue, I used a thin disk approximation. The spins are projected along a given line-of-sight direction in the box and the apparent axis ratio is assumed to be well-approximated by $q = |L_z|/|\mathbf{L}|$, where z is the line of sight direction. The orientation of the major axis of the projected ellipse is $\psi = \pi/2 - \arctan(L_y/L_x)$ so that the complex ellipticity can be written $e = (1-q)/(1+q)\exp(2i\psi)$ in cartesian coordinates. With this prescription, the projected correlation functions for a given projected separation θ can be estimated. The spins of blue

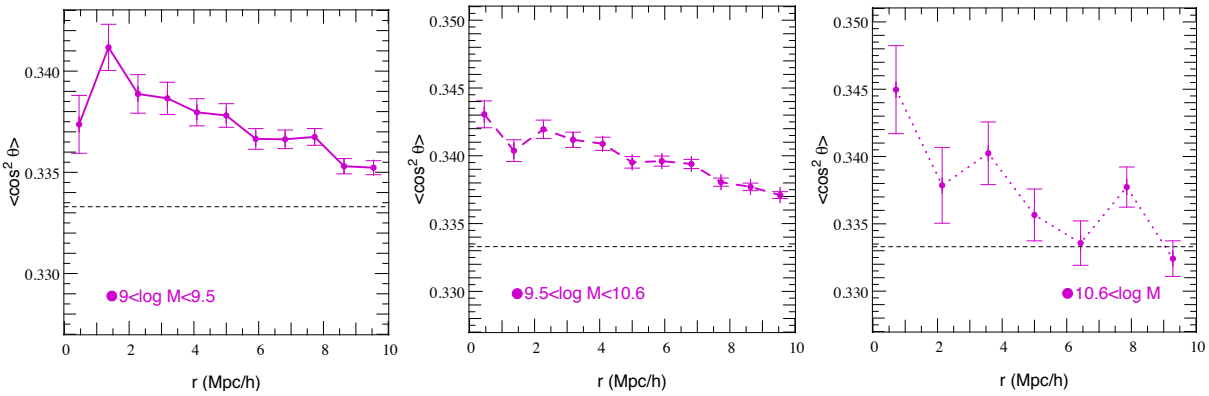


Figure 4.5: Mean square cosine of the angle between semi-minor axes as a function of the separation for different bins of mass as labelled. Measurements are done at redshift $z = 0.5$.

galaxies are shown to be correlated on scales about 10 arcminutes while (as expected from the 3D study) the signal for red galaxies is compatible with zero. Note that this signal is not contradictory with current observations as it is at a larger redshift.

A follow-up of this work is to investigate the redshift evolution of those intrinsic alignments together with a better estimation of the projected ellipticities using inertia tensors to define galaxy shapes instead of spins. Figure 4.5 shows the mean square cosine of the angle between semi-minor axes as a function of the separation for different bins of mass as labelled. A clear detection of alignment is found for all bins of mass and this correlation of minor axes pervades on distances as large as $\approx 10\ h^{-1}$ Mpc. The signal is very similar to what was obtained using the spins as expected since spin and minor axis are tightly correlated as shown on figure 4.6 which displays the cumulative PDF of the cosine of the angle between spin and semi-minor axis. The main difference between using the spins or the shapes reside in the larger mass bin. Indeed, in this case, galaxies are expected to be mainly ellipticals and therefore the spin – which is poorly defined – and the minor axis are less aligned.

The redshift evolution of the spin-spin correlation function is shown on figure 4.7 for three different redshifts: $z = 2$, 1.2 and 0.5. Correlations are increasing with time. Similar results are obtained for the shape-shape correlation function defined by $\eta(r)$ and inferred from the spins using an infinitely thin disk approximation (see Fig. 4.8).

The main worry in the study of spin-spin or shape-shape correlations using AMR simulations is grid locking. Indeed, as forces are evaluated on a grid, the spins and shapes of galaxies in such simulations are more likely to be aligned with the directions of the grid. This effect generates spurious correlations that are difficult to remove from the physical signal. Different strategies can be proposed to evaluate this effect. A simple randomization of galaxy positions is obviously not sufficient as the effect of grid-locking can be correlated on small scales: two neighbouring galaxies feel a similar force field and therefore tend to be similarly grid-locked. An alternative approach could be to select the galaxy sample (by stellar age, colour, mass, etc) that is the least prone to grid locking but this procedure significantly reduces the number of galaxies considered and therefore the statistics. Moreover, it is not clear how these selection cuts affect the intrinsic alignment measurement and its comparison to current observational constraints. However, it is possible to study another quantity – that is often used in the field of intrinsic alignments – the orientation-separation correlation as a function of comoving distance r,

$$\eta_r(r) = \langle |\hat{\mathbf{r}} \cdot \hat{\mathbf{e}}(\mathbf{x} + \mathbf{r})|^2 \rangle - 1/3\,, \tag{4.7}$$

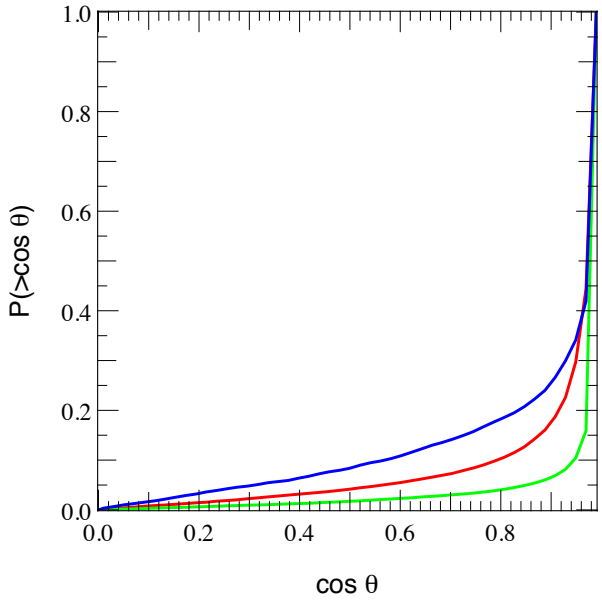


Figure 4.6: cumulative PDF of the cosine of the angle between spins and minor axes. Red is the low-mass bin, green the intermediate one and blue the high mass bin.

where $\hat{\mathbf{e}}$ is the unit eigenvector of the inertia tensor pointing in the direction of the minor axis and $\hat{\mathbf{r}}$ is the separation vector. A positive correlation indicates a tendency for the separation vector and the minor axis of a galaxy to be parallel, hence for the galaxy to be elongated tangentially with respect to another galaxy. A negative correlation corresponds to a preferential perpendicular orientation of the minor axis with respect to the separation vector, resulting in a net radial orientation of galaxy shapes around other galaxies. The advantage of this observable is that it is much less prone to grid locking as the distribution of galaxies in the simulation is not expected to be correlated with the grid.

Preliminary results on the correlation between galaxy's separations and shapes have recently been obtained (see paper 4.5). In this paper, we focus on the correlation between separation and shape of galaxies in the Horizon-AGN simulation at low redshift ($z \sim 0.5$), where observational constraints on the intrinsic alignment amplitude and scale-dependence are currently available. We find that spheroidal galaxies in the simulation show a tendency to be aligned radially towards overdensities in the dark matter density field and other spheroidals. The qualitative trend in agreement with observations, but the amplitude of the signal depends strongly on how shapes are measured and how galaxies are selected in the simulation. Disk galaxies show a tendency to be oriented tangentially around spheroidals. While this signal is suppressed in projection, we cannot guarantee that disk alignments can be safely ignored in future weak lensing surveys. We also characterize the systematics of galaxy shapes in the simulation and show that they can be safely neglected when measuring the correlation of the density field and galaxy ellipticities.

This work is a first step in the accurate modelling of intrinsic alignments and paves the way to future more realistic studies. One major improvement that could be done to get closer to observations would be, for instance, to do light-weighted measurements on the light cone and take into account observational systematics. This will be the subject of further investigations.

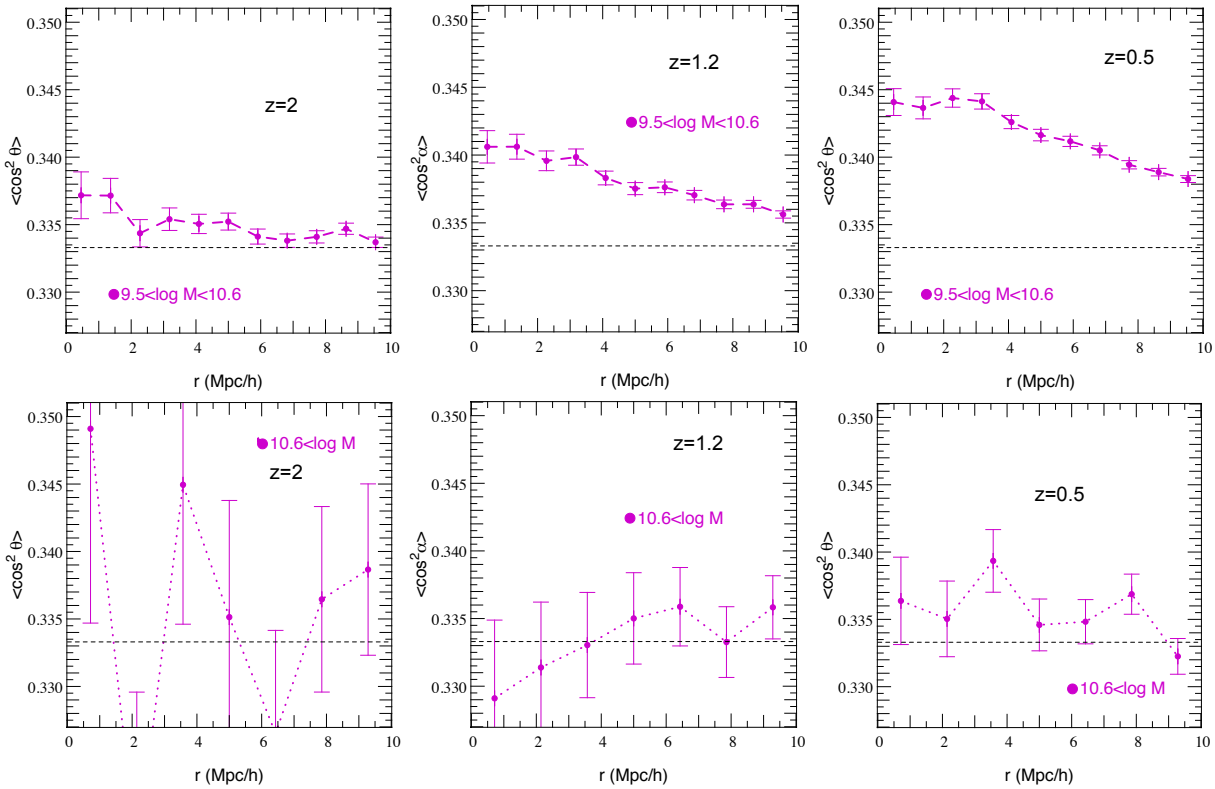


Figure 4.7: 3D spin-spin two-point correlation function of galaxy as a function of the comoving separation for a range of stellar masses: $3 \times 10^9 < M_s < 4 \times 10^{10}\,M_\odot$ at $z = 2$ (top left panel), $z = 1.2$ (top center panel), and $z = 0.5$ (top right panel). Bottom panels: same as top panels for stellar mass $M_s > 4 \times 10^{10}\,M_\odot$. Error bars represent the error on the mean. Here we choose to display the mean square cosine between two spins (separated by the comoving distance r) as the polarity is not relevant to weak lensing studies. For a uniform random distribution, the expectation is 1/3 (dashed line).

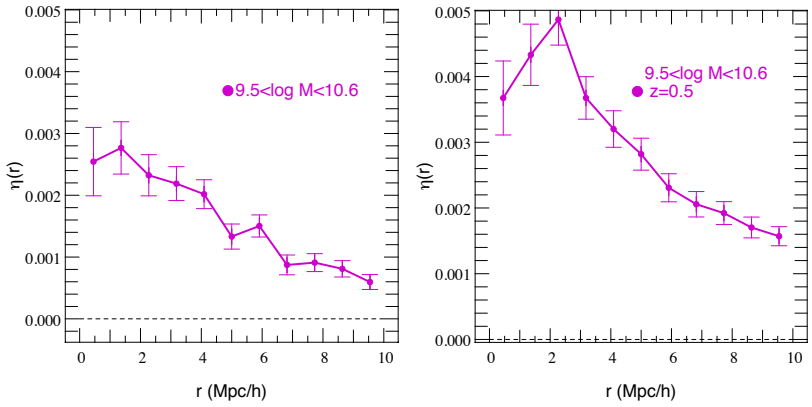


Figure 4.8: Correlation functions of the projected ellipticity $\eta(r)$ for middle-mass galaxies at $z = 1.2$(left panel) and 0.5 (right panel) as a function of the comoving 3D separation r

4.1.5 Conclusion

In short, we have shown here that dark matter halos and the galaxies they host are expected to be strongly correlated to the large-scale cosmic web. A particular emphasis was put on their shape and spin as they are crucial in the context of galaxy formation and intrinsic alignments. Those correlations are predicted coherently by numerical simulations and analytical calculations. The fact that tidal torque theory succeeds in explaining the correlations between spin and filaments seen in simulations is a great success. However it should be noted that those correlations are generated by the large-scale (linear!) dynamics which is why linear tidal torques are sufficient to describe them. Non-linearities on small scales necessarily dilute this large-scale coherence so that there is no reason to think that tidal torque theory is able to predict the low-redshift angular momentum of galaxies on a one-to-one basis – as already pointed out in numerous papers–. The individual history of each galaxies including mergers contribute significantly to the late-time evolution of the spin that will depart from tidal torque theory's predictions.

The intrinsic alignments of galaxies are therefore nicely understood as the result of the large-scale cosmic flows. However in order to be able to predict them with high accuracy, there is a real need to go beyond linear theory or semi-analytical models that can only provide a global picture. Hydrodynamical simulations seem to be the only way to predict the contamination by intrinsic alignments expected for future surveys as it is highly sensitive to baryonic physics, and therefore to any selection effects in the sample of galaxies (w.r.t galaxy properties such as their colour, mass, morphology), and to observational systematics. Only hydrodynamical simulations can model all those effects with sufficient accuracy. The worrisome counterpart of this statement is that those simulations have lots of parameters that encode our lack of knowledge of subgrid baryonic physics and a natural question that arises is therefore : do intrinsic alignments's prediction depend on the codes and subgrid recipes used in the simulation? What is the effect of using AMR vs SPH schemes? How much does AGN feedback play a role in aligning – or misaligning – galaxies? All those questions that are left unanswered for now represent a formidable field of investigation and should be crucial in the coming years.

All the ideas presented in this section are extensively described in four papers (three published articles I have signed as first author and one paper accepted for publication in MNRAS) that are copied below. Paper 4.2 shows the correlation between spin and filaments in the Horizon 4π simulation, paper 4.3 then exhibits the anisotropic tidal torque theory that takes into account the large-scale filamentary structure for spin acquisition. Finally, papers 4.4 and 4.5 propose an estimation of the level of intrinsic alignments in the state-of-the-art Horizon-AGN simulation.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **427,** 3320–3336 (2012) doi:10.1111/j.1365-2966.2012.21636.x

Connecting the cosmic web to the spin of dark haloes: implications for galaxy formation

Sandrine Codis,[1] Christophe Pichon,[1,2★] Julien Devriendt,[2] Adrianne Slyz,[2] Dmitry Pogosyan,[3] Yohan Dubois[1] and Thierry Sousbie[1]

[1]*Institut d'Astrophysique de Paris, 98 bis boulevard Arago, 75014 Paris, France*
[2]*Sub-department of Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH*
[3]*Department of Physics, University of Alberta, 11322-89 Avenue, Edmonton, Alberta T6G 2G7, Canada*

Accepted 2012 June 28. Received 2012 April 27; in original form 2012 January 27

ABSTRACT

We investigate the alignment of the spin of dark matter haloes relative (i) to the surrounding large-scale filamentary structure, and (ii) to the tidal tensor eigenvectors using the Horizon 4π dark matter simulation which resolves over 43 million dark matter haloes at redshift zero. We detect a clear mass transition: the spin of dark matter haloes above a critical mass $M_0^{\rm s} \approx 5(\pm1) \times 10^{12}\,{\rm M}_\odot$ tends to be perpendicular to the closest large-scale filament (with an excess probability of up to 12 per cent), and aligned with the intermediate axis of the tidal tensor (with an excess probability of up to 40 per cent), whereas the spin of low-mass haloes is more likely to be aligned with the closest filament (with an excess probability of up to 15 per cent). Furthermore, this critical mass is redshift-dependent, scaling as $M^{\rm s}_{\rm crit}(z) \approx M^{\rm s}_0(1+z)^{-\gamma_{\rm s}}$ with $\gamma_{\rm s} = 2.5 \pm 0.2$. A similar fit for the redshift evolution of the tidal tensor transition mass yields $M^{\rm t}_0 \approx 8(\pm2) \times 10^{12}\,{\rm M}_\odot$ and $\gamma_{\rm t} = 3 \pm 0.3$. This critical mass also varies weakly with the scale defining filaments.

We propose an interpretation of this signal in terms of large-scale cosmic flows. In this picture, most low-mass haloes are formed through the winding of flows embedded in misaligned walls; hence, they acquire a spin parallel to the axis of the resulting filaments forming at the intersection of these walls. On the other hand, more massive haloes are typically the products of later mergers along such filaments, and thus they acquire a spin perpendicular to this direction when their orbital angular momentum is converted into spin. We show that this scenario is consistent with both measured excess probabilities of alignment with respect to the eigendirections of the tidal tensor, and halo merger histories. On a more qualitative level, it also seems compatible with 3D visualization of the structure of the cosmic web as traced by 'smoothed' dark matter simulations or gas tracer particles. Finally, it provides extra support to the disc-forming paradigm presented by Pichon et al. as it extends it by characterizing the geometry of secondary infall at high redshift.

Key words: methods: numerical – galaxies: formation – galaxies: haloes – large-scale structure of Universe.

1 INTRODUCTION

Over the past decades, numerical simulations and large redshift surveys have highlighted the large-scale structure (LSS) of our Universe, a cosmic web formed by voids, sheets, elongated filaments and clusters at their nodes (Bond, Kofman & Pogosyan 1996). This structure is believed to be the result of the linear growth of primordial Gaussian fluctuations in a nearly homogeneous density field followed by the non-linear collapse of overdense regions into dark matter (DM) haloes which then accrete mass and merge as described by the hierarchical model. The current paradigm of galaxy formation states that collapsing protogalaxies acquire their spin (i.e. their angular momentum) by tidal torquing because of a misalignment between their inertia tensor and the local gravitational tidal tensor at the time of maximum expansion; this is the basis of the so-called tidal torque theory (hereafter TTT, Hoyle 1949; Peebles 1969; Doroshkevich 1970; White 1984; Catelan & Theuns 1996; Crittenden et al. 2001; Schäfer 2009, for a recent review). According to this theory, the spin direction should initially be correlated with

★E-mail: pichon@iap.fr

the principal axes of the local tidal tensor, defined as the traceless part of the Hessian matrix of the gravitational potential field. One therefore expects to detect correlations between actual galactic angular momenta and LSSs if non-linear processes have not modified their direction.

For many years, both observers and theorists have thus endeavoured to detect these correlations in real surveys and cosmological N-body simulations. None the less, the results remain in part contradictory because of the lack in resolution together with the difficulty of properly defining large-scale filamentary structures.

For instance, using N-body simulations Hahn et al. (2007a), Sousbie, Colombi & Pichon (2009) and Zhang et al. (2009) found that halo spins are preferentially oriented perpendicularly to the filaments whatever their mass, with Faltenbacher et al. (2002) measuring a random distribution of the spin in the plane perpendicular to the filaments, while Hatton & Ninin (2001) claimed an alignment between spin and filament. More recently a consensus seemed to have emerged when several works (Bailin & Steinmetz 2005; Aragón-Calvo et al. 2007; Hahn et al. 2007b; Paz, Stasyszyn & Padilla 2008) reported that LSSs – filaments and sheets – influenced the direction of the angular momenta of DM haloes in a way originally predicted by Sugerman, Summers & Kamionkowski (2000) and Lee & Pen (2000). These studies pointed towards the detection of a mass-dependent orientation of the spin, arguing for the first time that the spin of high-mass haloes tends to be perpendicular to their host filament, whereas low-mass haloes have a spin preferentially aligned with their host filament. Nevertheless, the detected correlation remains weak and noisy and no full explanation for these findings was highlighted except, for example, Bailin & Steinmetz (2005) who suggested that the spin direction of cluster and group mass haloes (as opposed to galaxy mass haloes) could come from mergers along the filaments.

However, Hahn, Teyssier & Carollo (2010) repeated this measurement in a cosmological hydrodynamical resimulation of one large-scale cosmic filament and found a different result: the spin of high-mass haloes is aligned with the filament; the spin of low-mass haloes is along the intermediate eigendirection of the tidal tensor in low-density regions and along the third eigendirection (i.e. neither the intermediate direction nor the filament's) in higher density regions at higher redshift ($z = 1$); finally, no signal exists for high-density regions at low redshift.

Beyond numerical simulations, Lee & Erdogdu (2007), Trujillo, Carretero & Patiri (2006), Navarro, Abadi & Steinmetz (2004), Flin & Godlowski (1990, 1986) and Godłowski & Flin (2010) found correlations in observations between the rotational axis of galaxies and the surrounding LSSs (e.g voids and local tidal shears) unlike Dekel (1985) who did not find any correlation. To be more specific, Flin & Godlowski (1986, 1990) first discovered that the spin of galaxies is not isotropically oriented with respect to the Local Supercluster plane but more likely to be aligned with it; Navarro et al. (2004) confirmed this observation; Trujillo et al. (2006) found that in SDSS and 2dFGRS, the rotational axis of the spiral galaxies located in the walls surrounding voids lie preferentially in the plane of these voids; Lee & Erdogdu (2007) analysed the galaxies of the 2MASS Redshift Survey and found correlations between their spin and the local tidal tensor and Godłowski & Flin (2010) focused on the galaxy groups in the Local Supercluster and found correlations in their orientation, suggesting that the two brightest galaxies and then the galaxy groups were hierarchically formed in the same filament with their major vector aligned with this host filament. The results of Navarro et al. (2004) and Trujillo et al. (2006) support the predictions of TTT (Lee & Pen 2000) namely that, assuming that the inertia and tidal tensors are uncorrelated, galaxies' spin should be preferentially aligned with the intermediate eigendirection of the tidal tensor (in particular in the plane of the voids). Recently, however, Slosar & White (2009) claimed that in contrast to previous studies, they found no departure from randomness in SDSS while studying the orientation of the galaxy spin with regard to the voids in which they are located. The method used in the latter study has been improved by Varela et al. (2012) who used the SDSS (DR7) and morphological classifications from the Galaxy Zoo project and found that galaxy discs are more likely to lie in the plane of voids, that is, their spin tends to be perpendicular to the void they are located in, which seems to be in disagreement with Lee & Erdogdu (2007), Trujillo et al. (2006) and Navarro et al. (2004).

In short, even if one can claim that a trend is slowly emerging, quantitative evidence of spin alignment with the filaments and tidal tensor eigendirections remains at this stage weak and somewhat inconclusive. Hence, in this paper, we propose to revisit the issue and quantify the alignment between the spin of DM haloes and the filamentary pattern in which they are embedded (together with the alignment between the spin and the tidal tensor principal axes) using a very efficient topological tool, the Skeleton (Sousbie et al. 2009). This tool provides a robust and mathematically well defined reconstruction of the cosmic web filaments. We apply it to the Horizon 4π simulation (Teyssier et al. 2009), a $2\,h^{-1}$ Gpc on a side cubic volume of the Universe containing over 67 billion DM particles which provides an unprecedented catalogue of 43 million DM haloes with masses $>2 \times 10^{11}\,\mathrm{M}_{\odot}$. We then interpret our results in the framework of the dynamics of large-scale cosmic flows.

Section 2 briefly presents the Horizon 4π simulation and the topological tool implemented to identify the loci and orientation of filaments. It then reports the correlations detected between the orientation of the spin of DM haloes and filaments and its redshift evolution. Section 3 is devoted to the physical processes that induce these correlations. It also illustrates them using DM halo merging histories, smoothed DM simulations and hydrodynamical simulations. Section 4 provides conclusions and discusses prospects for our understanding of galaxy formation within its cosmic environment. Appendix A gives the correlations measured between spin directions and tidal eigendirections, which are in agreement with the cosmic dynamics arguments of Section 2. Appendix B presents a visual quantitative estimation of the spin of the circumgalactic medium. Appendix C sums up all the tests we have performed to assess the robustness of the correlations presented in Section 2. Finally, Appendix D presents the dependence of the transition masses on the smoothing length and the non-linear mass as a function of redshift.

2 SPIN–FILAMENT CORRELATIONS

Let us first account for the robust correlation between the DM halo's spin and the orientation of the filaments of the cosmic web.

2.1 Virtual data sets

This study uses the Horizon 4π N-body simulation (Teyssier et al. 2009) which contains 4096^3 DM particles distributed in a $2\,h^{-1}$ Gpc periodic box to investigate the spin alignment of DM haloes relative to their LSS environment. This simulation is characterized by the following Λ cold dark matter (ΛCDM) cosmology: $\Omega_{\mathrm{m}} = 0.24$, $\Omega_{\Lambda} = 0.76$, $n = 0.958$, $H_0 = 73\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\sigma_8 = 0.77$ within 1 standard deviation of *WMAP3* results (Spergel et al. 2003).

These initial conditions were evolved non-linearly down to redshift zero using the adaptive mesh refinement code RAMSES (Teyssier 2002), on a 4096^3 grid. The motion of the particles was followed with a multigrid Particle-Mesh Poisson solver using a Cloud-In-Cell interpolation algorithm to assign these particles to the grid (the refinement strategy of 40 particles as a threshold for refinement allowed us to reach a constant physical resolution of 10 kpc; refer to the above-mentioned two references).

The Friend-of-Friend (FoF) algorithm (Huchra & Geller 1982) was used over 18^3 overlapping subsets of the simulation with a linking length of 0.2 times the mean interparticular distance to define DM haloes. In this work, we only consider haloes with more than 40 particles, which corresponds to a minimum halo mass of $3 \times 10^{11}\,\mathrm{M}_\odot$ (the particle mass is $7.7 \times 10^9\,\mathrm{M}_\odot$). The mass dynamical range of this simulation spans about five decades. Overall, 43 million haloes were detected at redshift zero (see Fig. 1). This simulation was complemented by smaller (1024^3 particles, box size $200\,h^{-1}$ Mpc and several 256^3 particles, box size $50\,h^{-1}$ Mpc, leading to a particle mass of $6.2 \times 10^8\,\mathrm{M}_\odot$) DM-only simulations to address resolution issues (see Appendix C) and interpret the redshift evolution of the signal (see Section 2.3).

Several topological techniques (Barrow, Bhavsar & Sonoda 1985; Stoica et al. 2005; Novikov, Colombi & Doré 2006) have been proposed to identify the complex cosmic network made of large voids, sheet-like structures and elongated filaments. These techniques rely on giving a mathematical definition (and a detection algorithm) of the filamentary pattern that our eye easily detects in the simulations. One recent criterion to classify structures as clusters, filaments, sheets or voids uses the number of positive eigenvalues of the Hessian matrix of the density or potential fields (Aragón-Calvo et al. 2007; Hahn et al. 2007b; Forero-Romero et al. 2009). Another interesting approach was followed by Platen, van de Weygaert & Jones (2007, 2008) who used Watershed transforms to identify voids, filaments and walls. More recently, González & Padilla (2010) introduced a method which relies on positions and masses of DM haloes, and Bond, Strauss & Cen (2010) used Hessian eigendirections to detect filaments. In this paper, we use the (global) 3D Skeleton introduced by Sousbie et al. (2009). The underlying algorithm is based on Morse theory and defines the Skeleton as a set of critical lines joining the maxima of the density field through saddle points following the gradient. In practice, Sousbie et al. (2009) define the peak and void patches of the density field as a set of points converging to a specific local maximum/minimum while following the field lines in the direction/opposite direction of the gradient. The Skeleton is then a set of intersections of the void patches, that is, a subset of critical lines connecting the saddle points and the local maxima of a density field and parallel to the gradient of the field.

For this work, the ∼70 billion particles of the Horizon 4π simulation were sampled on a 2048^3 Cartesian grid and the density field was smoothed over 5σ using `mpsmooth` (Prunet et al. 2008), corresponding to a scale of $5\,h^{-1}$ Mpc and a mass of $1.9 \times 10^{14}\,\mathrm{M}_\odot$. Hence, we are focusing on the LSS of the cosmic web. The corresponding cube was then divided into 6^3 overlapping subcubes (with a buffer zone of 100 voxels in each direction, large enough to cover the largest peak patches of the simulation, see Sousbie et al. 2009), and the Skeleton was computed for each of these subcubes. It was then reconnected across the entire simulation volume to produce a catalogue of segments which locally define the direction of the Skeleton. This Skeleton is shown in Fig. 2. Note that this Skeleton (i.e. what we will call filaments in the rest of this paper) depends on the choice we made for the smoothing length ($5\,h^{-1}$ Mpc). Appendix D1 investigates the effect of probing smaller smoothing scales on other sets of simulations.

The hydrodynamical simulations used in this paper are described in Appendix B.

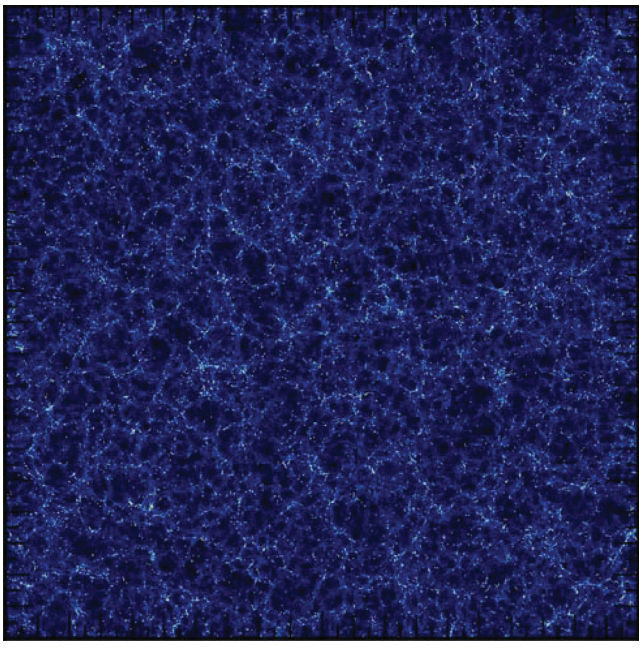


Figure 1. An 80 h^{-1} Mpc slice of the Horizon 4π simulation at redshift zero. The box size is $2\,h^{-1}$ Gpc across. On top of the DM log-density (colour coded in levels of blue, from dark to light), all haloes in that slice whose mass is larger than $3 \times 10^{13}\,\mathrm{M}_\odot$ are shown in the yellow dots. As expected, these haloes fall on top of the filamentary structure of the cosmic web.

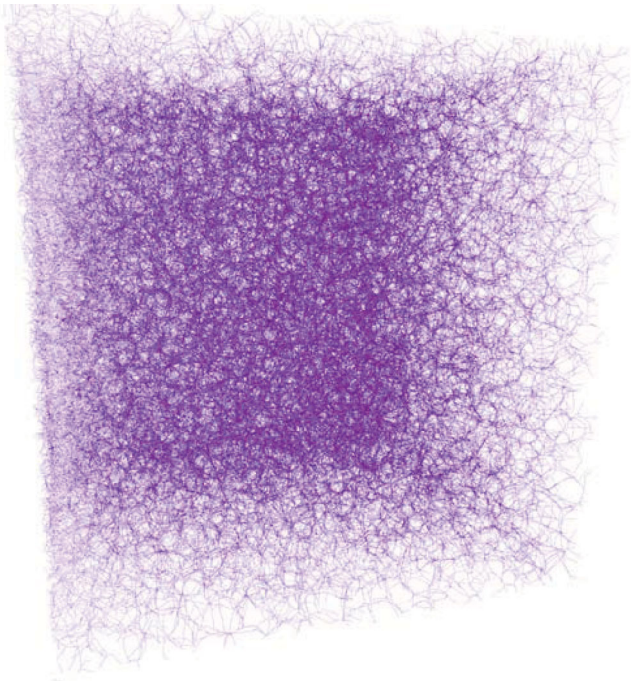


Figure 2. A 3D view of the Skeleton of the Horizon 4π simulation measured from the DM distribution. The size of the box is $2\,h^{-1}$ Gpc. This paper analyses the relative orientation of the spin of the 43 million dark haloes relative to the cosmic web.

2.2 Correlations between spin and filament axis

In order to study the alignment between the spin of haloes and the filamentary features of the cosmic web, we compute the Skeleton of the LSS for the density field smoothed with the above quoted Gaussian scale $R = 5\,h^{-1}$ Mpc which corresponds to $\sigma(R) = 0.66$. Thus, we are considering the filaments that are mildly non-linear LSSs at cluster scales.

In this paper, the spin of a given halo is defined as the sum over its particles denoted by i: $m_{\rm p}\sum_i(\boldsymbol{r}_i - \bar{\boldsymbol{r}}) \times (\boldsymbol{v}_i - \bar{\boldsymbol{v}})$, where $\bar{\boldsymbol{r}}$ is the centre of mass of the FoF halo and $\bar{\boldsymbol{v}}$ is its mean velocity. We search for the five DM haloes (regardless of their mass) closest to *each* filament segment (see Appendix C for alternative choices). We then measure the angle between the angular momentum of these haloes and the direction of the filament segment and estimate the probability distribution function (PDF) of the absolute value of the cosine of this angle; this PDF, $1 + \xi$, measures the excess probability of alignment between the halo spin and the direction of the filament (note in particular that it is normalized for $\cos\theta$ between 0 and 1; for aesthetic purpose only, data are symmetrically plotted for $\cos\theta$ between -1 and 1; Appendix C briefly discusses the associated biases). The data are split by halo masses ranging from galactic to cluster masses and are displayed in Fig. 3, the main result of this paper.

A clear signal is detected. The orientation of the halo spin depends on the local anisotropy of the cosmic web, *and* on the DM halo mass: the spin of DM haloes is preferably perpendicular to their host filament at high mass (with an excess probability reaching 12 per cent), but turns into being aligned with the nearest filament direction at lower masses (with an excess probability of 15 per cent). This 'phase transition' is found to occur at $M^{\rm s}_{\rm crit}(z = 0) \simeq 4(\pm1) \times 10^{12}\,{\rm M}_\odot$, where $M^{\rm s}_{\rm crit}$ is defined as the halo mass for which $\langle\cos\theta\rangle = 0.5$. Fig. 4 shows an example segment of the large-scale filamentary network together with the orientation of spins of massive haloes that graphically demonstrates for them the effect of spin–filament anti-alignment. Several sanity checks have been carried out to assess the robustness of this signal and are summed up in Appendix C.

These measurements of the spin–filament correlation trend confirm the previous results obtained by Bailin & Steinmetz (2005), Aragón-Calvo et al. (2007), Hahn et al. (2007b) and Paz et al. (2008) with significantly improved statistics which allow us to quantify the mass transition.

2.3 Redshift dependence of the transition mass

The redshift dependence of the transition mass was then investigated on a set of smaller ΛCDM simulations (1024^3 particles in $200\,h^{-1}$ Mpc periodic boxes and 256^3 particles in $50\,h^{-1}$ Mpc periodic boxes). At high redshift we define the filamentary structure at the smoothing scale, $R(z)$, chosen to maintain the same level of non-linearity as we had at redshift zero for $R_0 = 5\,h^{-1}$ Mpc. Thus, the smoothing scale $R(z)$ is obtained from the implicit condition $\sigma^2(R(z), z) = \sigma^2(R_0, 0)$ (see Appendix D2). The halo spins continue to exhibit a transition from alignment with the nearest filament at low mass to anti-alignment at high mass. The critical mass for the transition, $M^{\rm s}_{\rm crit}(z)$, is found to decrease with redshift as a power law

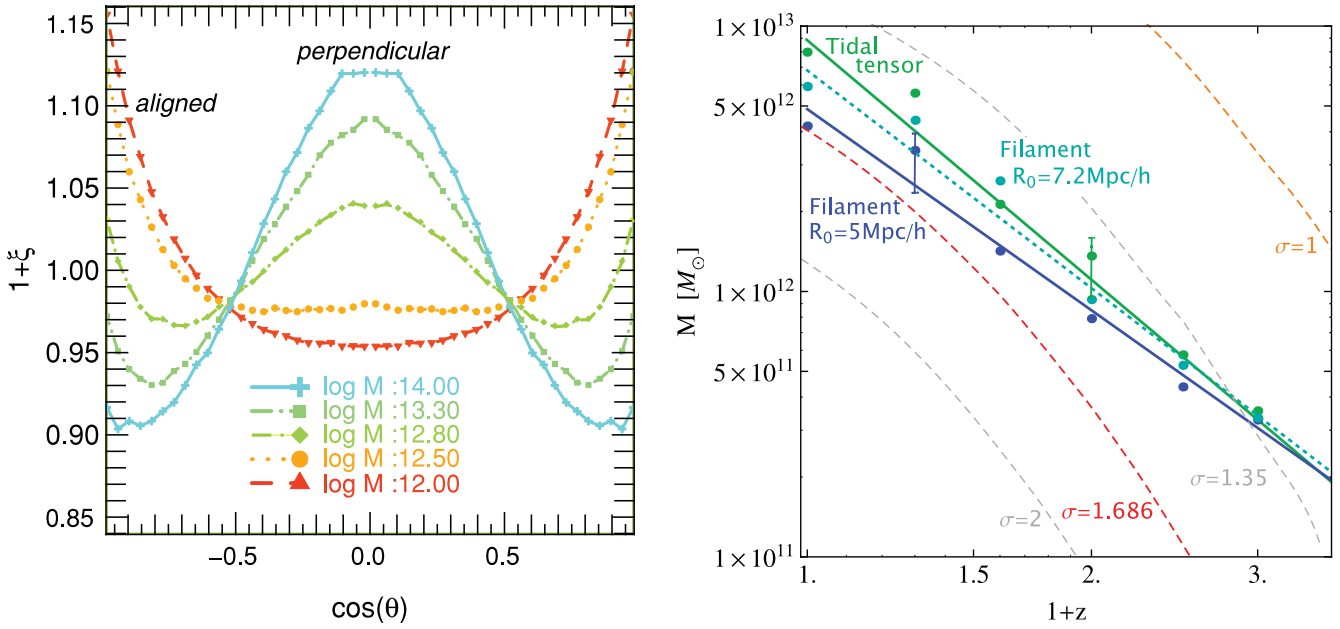


Figure 3. Left-hand panel: excess probability of alignment between the spin and the direction of the closest filament as measured from the 43 millions haloes of the Horizon 4π simulation at redshift zero. Different colours correspond to different mass bins from 10^{12}(red) to $10^{14}\,{\rm M}_\odot$ (blue) as labelled. Thanks to the very large number of haloes in each mass bin, the excess probability is quite well sampled and displays a clear departure from a uniform distribution. A transition mass is detected at $M^{\rm s}_0 = M^{\rm s}_{\rm crit}(z = 0) \simeq 5(\pm1) \times 10^{12}\,{\rm M}_\odot$: for haloes with $M > M^{\rm s}_0$, the spin is more likely to be perpendicular to their host filament, whereas for haloes with $M < M^{\rm s}_0$, the spin tends to be aligned with the closest filament. Right-hand panel: redshift evolution of the filament transition mass (in blue) and the tidal tensor transition mass (in green) derived from the $200\,h^{-1}$ Mpc ΛCDM simulations as discussed in the main text. The (cyan) dotted line represents the spin–filament mass transition for a larger smoothing length ($7.2\,h^{-1}$ Mpc). The displayed error bar is estimated as one-third of the bin mass. The dashed lines correspond to the non-linear masses (for a top-hat filter, see Appendix D2) at different σ, in particular $\sigma = 1$ (orange) and $\sigma = 1.686$ (red). The redshift evolution of the transition masses is in qualitative agreement with that of $M_{\rm NL}(\sigma \lesssim 1.686)$ though they seem to remain close to power laws throughout the explored range of redshift.

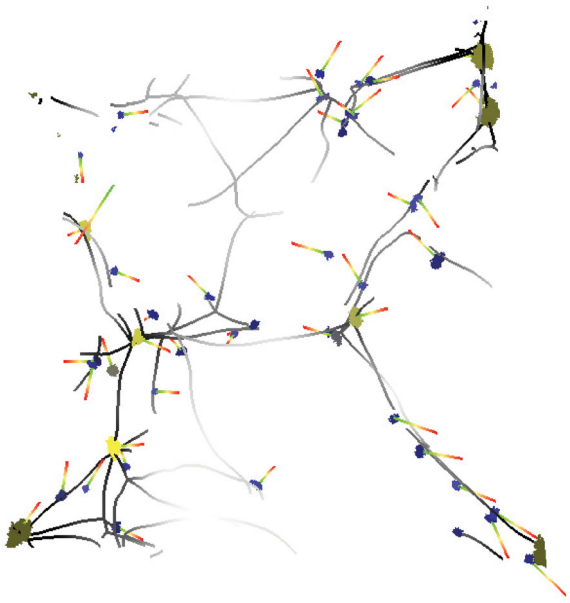


Figure 4. A very small subset of Skeletons at different redshifts extracted from the Horizon 4π simulation (see Fig. 2), together with unit vectors showing the orientation of the spin of the corresponding DM halo with a mass above the transition mass. The spin is indeed perpendicular to the filament for these massive haloes.

of z. Namely,

$$M^{\rm s}_{\rm crit} \approx M^{\rm s}_0(1+z)^{-\gamma_{\rm s}}, \quad \gamma_{\rm s} = 2.5 \pm 0.2, \quad M^{\rm s}_0 \simeq 5\times10^{12}\,{\rm M}_\odot. \tag{1}$$

This result is presented in Fig. 3 over the studied range $z = 0$–4. Measuring the dependence of the $z = 0$ transition mass M_0 on scale R_0 (see Fig. D1), we find some weak scaling, $M^{\rm s}_0 \propto R_0^{0.8}$. Note that this dependence significantly depends on redshift.

The existence and redshift dependence of the transition mass in spin–structure alignment is supported by studying the halo's spin direction relative to the orientation of the large-scale gravitational tidal tensor. The details are given in Appendix A, where we find that the more massive haloes are preferably aligned with the intermediate principal axis of the tidal tensor, while smaller haloes show a positive alignment with the minor axis (which near a filament is the direction in which the filament extends). In this approach, our measurements give for the transition mass

$$M^{\rm t}_{\rm crit} \approx M^{\rm t}_0(1+z)^{-\gamma_{\rm t}}, \quad \gamma_{\rm t} = 3 \pm 0.3, \quad M^{\rm t}_0 \simeq 8\times10^{12}\,{\rm M}_\odot, \tag{2}$$

in good agreement with the Skeleton results, equation (1). The somewhat larger amplitude of $M^{\rm t}_0$ with respect to $M^{\rm s}_0$ can be explained by noticing that the tidal tensor associated with the gravitational potential smoothed on a scale R_0 effectively probes larger scales than the smoothed density field itself. Fig. 3 shows that if we boost the smoothing scale used to define the Skeleton to $R_0 = 7.2\,h^{-1}$ Mpc, $M^{\rm s}_{\rm crit}$ and $M^{\rm t}_{\rm crit}$ will match very closely (see also Appendix D1).

In Fig. 3, we also compare $M_{\rm crit}(z)$ to the redshift evolution of the mass scale $M_{\rm NL}(z)$ that corresponds to the fixed $\sigma(R, z)$ (defined in Appendix D2). Several values of the variance are of interest to track. One is $\sigma(R, z) = 1$ which formally defines the scale of non-linearity. Another is $\sigma(R, z) = 1.686$ which corresponds to the characteristic mass scale, $M_\star$, of collapsed gravitationally bound haloes at redshift z in the spherical top-hat model. Even though it is clear from Fig. 3 that both transition masses $M^{\rm s}_{\rm crit}$ and $M^{\rm t}_{\rm crit}$ qualitatively match a non-linear mass evolution with $\sigma \lesssim 1.686$ at low redshifts, at high z they still follow a power-law behaviour, while $M_{\rm NL}$ steepens as it probes the steepening power spectrum at ever shorter scales. Thus, at high redshifts, the positive alignment between the halo's spins and the nearby filaments extends to masses that are effectively higher, in terms of the corresponding characteristic non-linear mass. Although we do not have the full quantitative explanation for this effect, it may be related to the fact that the filaments at high z are more pronounced due to a steeper power spectrum, and are correlated with the shear of the surrounding flow more robustly. Note that the detection of haloes and filaments at these redshifts may be a concern for these intermediate-resolution simulations. Whilst we defer a detailed quantitative understanding of the redshift evolution of the mass transition, the rest of this paper is devoted to *explaining* the origin of these mass transitions.

3 SPIN INDUCED BY LSS DYNAMICS

The measurements of Section 2.2 strongly suggest that the spin direction of DM haloes is connected to the cosmic web. Indeed, these dark haloes are embedded in large-scale cosmic flows induced by the successive formation of walls, filaments and clusters: matter escapes from the voids to the walls then to the filaments before flowing along the latter direction towards the nodes (Zel'Dovich 1970; Bond et al. 1996; Pichon et al. 2011). In this framework, let us now argue that the first generation of haloes are formed during the *winding of walls around filaments*, which provides them with a spin parallel to this direction (and whose amplitude is proportional to the relative impact parameter of the two walls). Conversely, the later generations of haloes form by mergers along the filaments, that is, in the direction parallel to the mean flow (as was first pointed out by Sousbie et al. 2008) and therefore acquire a spin perpendicular to the filaments (if the orbital angular momentum which is converted into spin during the merger dominates). In this scenario, the transition in mass measured in Fig. 3 in fact reflects a trend in merging generation.[1] Note that this behaviour occurs on several scales simultaneously; this multiscale signal is probed by varying the smoothing scale used to define the filaments in Appendix D1.

As a first check for this hypothesis, the typical distance of DM haloes from filaments is computed as a function of their mass. The more massive haloes are typically found closer to the filaments (on average at 0.7 h^{-1} Mpc) than low-mass haloes (found on average at 2 h^{-1} Mpc from the core of the filaments), which is consistent with our assumption because it implies that high-mass haloes have reached the centre of the filaments, where they are more likely to merge in the direction of the flow. Conversely, low-mass haloes (for which large-scale dynamics have been frozen in at an early stage) are found farther from the core of the filaments where they are less likely to merge. This is qualitatively consistent with Fig. 1, which shows the distribution of massive haloes within a slice of the simulation.

3.1 Winding up of DM flows

Let us have a look at Fig. 5, which displays the temporal evolution of DM particle trajectories in the vicinity of a DM filament. This simulation has the special feature that the small-scale modes were

[1] Technically, it was not possible to build merging trees a posteriori on the Horizon 4π simulation as it would have required too many snapshots and thus too much disc space; so we did carry out the tests described in Section 3.2 on a smaller simulation to confirm the relevance of merger generation as the key parameter.

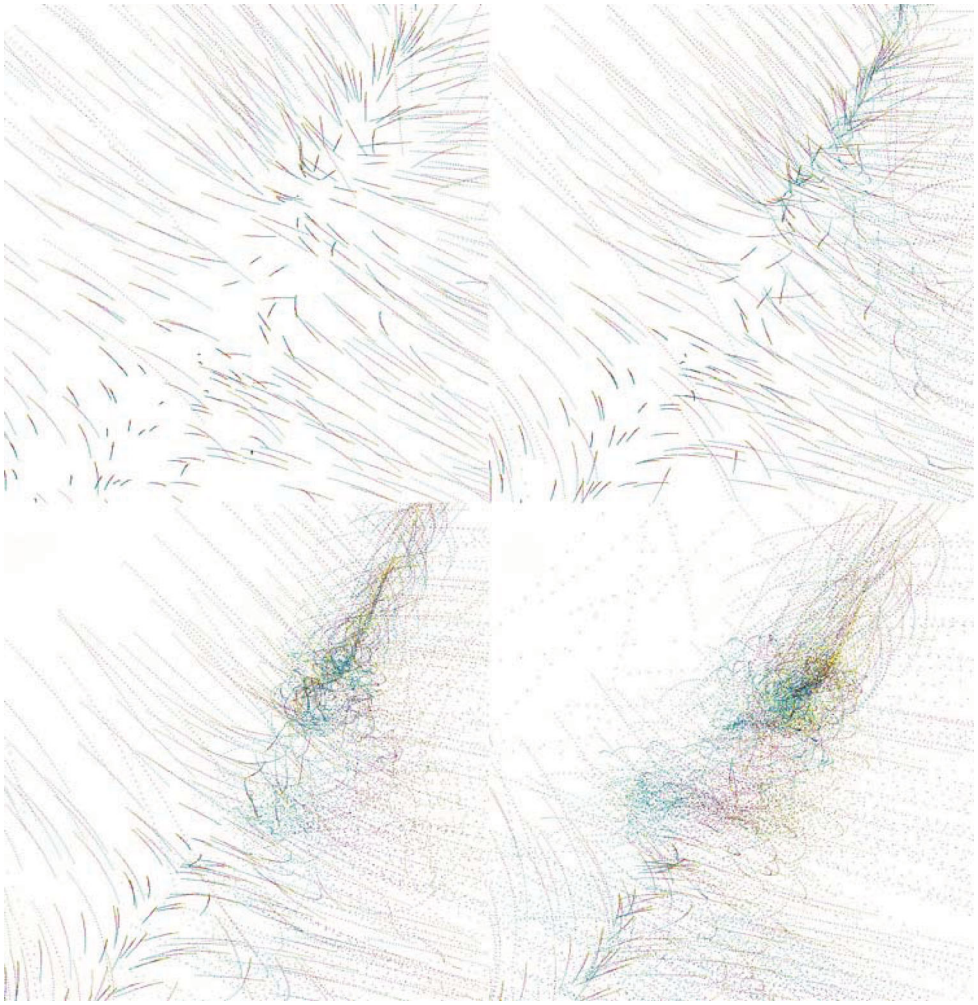


Figure 5. Trajectory of DM particles, colour coded from yellow (early) to blue (late) via dark red as a function of time. The four panels (from the top to bottom and left-hand to right-hand side) correspond to different stages of the winding of two walls around a north-eastern-oriented filament. Once DM has joined the filament, it heads towards the bottom left-hand part of the panel in the direction of a more massive node of the cosmic network. This process is best seen dynamically at http://www.iap.fr/users/pichon/spin/.

erased from the initial conditions in order to facilitate visualization of the large-scale flow (see Pichon et al. 2011 for details). We refer to this simulation as a 'smoothed' simulation. Here most of the flow is in fact embedded in a large wall parallel to the plane of the figure. The dots with different colours (from yellow to blue via dark red) correspond to the position of the same DM particle at successive time-steps, and thus allow us to visually follow the trajectory of DM particles. The top left-hand panel corresponds to a snapshot somewhat before the flow has significantly shell crossed around the north-eastern filament. The DM particles are plunging towards their filament, while flowing within the two walls. In the top right-hand panel, some level of shell crossing has occurred in the north-eastern part of the filament, and the corresponding particles have started inflecting their trajectories to wind up around the locus of that filament. Since these particles typically have a non-zero impact parameter relative to the centre of mass, as they wind up, they convert their orbital motion into spin while generating a virialized structure. Later on (bottom left-hand panel) this structure sinks along the north-eastern filament towards a more massive clump (off field). Meanwhile, the process of DM winding from the walls around the main filament continues, and feeds (as a yellow trail) the DM halo along its current spin axis (which is aligned with the axis of the filament). Finally, in the bottom right-hand panel, another such halo has formed farther down the filaments, and we can anticipate that their upcoming merger will lead to a structure whose spin's direction will be a mixture of their initial spin and the spin perpendicular to their relative orbital plane. In Section 3.3 below, we will revisit this scenario using hydrodynamical simulations, which will allow us to visually identify the spin of forming galaxies.

3.2 The progenitors of dark haloes via merging trees

In order to understand the previously described mass transition and its redshift evolution, we used the code TreeMaker (Tweed et al. 2009) to track down the progenitors (both dark haloes and unresolved flows) of given haloes in conjunction with their spin orientation relative to the nearest filament. TreeMaker involves two steps: first, haloes are identified using a halo finder – in our case a FoF algorithm (Huchra & Geller 1982; Zeldovich, Einasto & Shandarin 1982; Davis et al. 1985), while the properties of these structures (mass, angular momentum, etc.) are measured. As the second step, the individual DM particles which belong to each halo are tracked back in time so as to build a merging history tree which regroups all of its progenitors and their properties as a function of

time. A relatively large number of merging trees were computed and the direction of the spin of the progenitors relative to their host filament was calculated.

After visual inspection of a subset of those merging trees, it was found that: (i) the high-mass haloes (i.e. above the critical mass) with a spin perpendicular to their filament tend to have a similar history: they often acquire a significant amount of mass via a major merger, which is accompanied by a significant spin adjustment from a direction initially aligned with their filament to a direction mostly perpendicular to their filament (and aligned with $\boldsymbol{e}_2$, see Appendix A); and (ii) in contrast, low-mass haloes (i.e. below the critical mass) are not the result of major mergers; often no mergers at all are found at the mass resolution of the N-body simulations; those who have a spin parallel to the filament seem to have acquired this spin direction at a time (the so-called formation time) when they have acquired most of their mass by diffuse accretion.

This behaviour is quite generic as we observed it for a few tens of randomly chosen haloes. It is illustrated in Fig. 6 where merging trees of two different haloes (extracted from one of the 50 h^{-1} Mpc ΛCDM simulations) are shown: the evolution of the angle between the closest filament and the spin of a given halo is plotted as a function of redshift; the colours encode the fraction of mass of the progenitor with respect to the final halo mass. The right-hand panel corresponds to a low-mass halo ($2 \times 10^{11}\,\mathrm{M}_\odot$ at redshift zero corresponding to more than 300 particles) which forms at redshift $z \simeq 1.5$ (when it has already acquired more than one half of its mass) and suddenly acquires a spin parallel to its closest filament at a high redshift ($\simeq$2) close to its formation time. This halo does not undergo any significant merger afterwards. Conversely, the left-hand panel provides the merging tree of a high-mass halo ($8 \times 10^{12}\,\mathrm{M}_\odot$ at redshift zero) which forms at lower redshift ($\simeq$0.4) as the result of a major merger between two less massive haloes. This event corresponds exactly to the time when it acquires a spin perpendicular to its closest filament. What is striking here is the clear flip of the spin direction: the two progenitors have a spin aligned with the filament (this spin is acquired at higher redshift $\simeq$1.5) and their merger makes the spin of the resulting halo become perpendicular to it. These two examples of trees are characteristic of how haloes below and above the critical mass $M^{\mathrm{s}}_{\mathrm{crit}}$ form and acquire spin. Note that we actually observe a large dispersion of the histories around this mean behaviour.

To check the statistical robustness of this scenario for low-mass objects, let us identify a preferred plane of motion at formation time. The following simple test was implemented: a set of low-mass haloes ending with a spin parallel to their closest filament are randomly chosen and only haloes for which a time of significant accretion (i.e. their formation time) can be determined are retained (this represents at least one-third of our sample). Their particles are traced back one time-step before their formation time to quantify the relative orientation of their velocities compared to the filament's direction. The excess probability distribution of alignment, $1 + \xi_V$, is shown in Fig. 7 and shows that their velocities (before formation) are more likely to lie perpendicular to the filament (in particular it is found that $\langle \cos\theta_V \rangle \simeq 0.47$), which is in agreement with the scenario (see also Fig. A1).

In order to assess the statistical relevance of the scenario for massive haloes (i.e. above the critical mass at redshift zero), merging trees can also be used to determine their most recent merger. The mean cosine of the angle between these haloes and their host filament just after merging is computed and compared to the mean cosine of the angle between the closest filaments and the spin of the progenitors just before merging. This test yields $\langle \cos\theta \rangle \simeq 0.51$ before and 0.47 after the last merger, which is fully consistent in amplitude with the mean cosines found for the Horizon 4π simulation [for which, for instance, $\langle \cos\theta \rangle \simeq 0.510$ (0.479) for $M \simeq 10^{12}\,\mathrm{M}_\odot$ ($M \simeq 10^{13}\,\mathrm{M}_\odot$)]. The 1$\sigma$ error on the mean is about $\pm$0.02, given the size of this sample (only $\simeq$ 200 haloes massive enough in this simulation). This statistical test is in agreement with the fact that massive haloes acquire a spin perpendicular to their host filament because of mergers.

Altogether, we are now able to reconstruct the history of spin acquisition by DM haloes following the large-scale dynamics. The

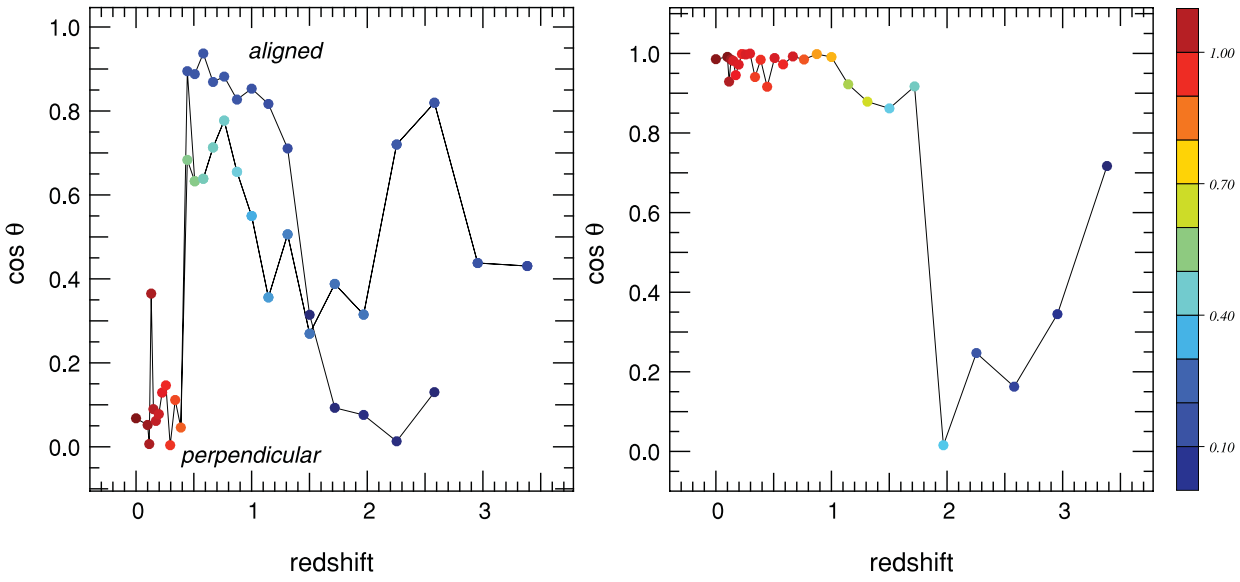


Figure 6. Left-hand panel: merging tree of a high-mass halo ($8 \times 10^{12}\,\mathrm{M}_\odot$ at redshift zero). The various colours correspond to different mass fractions (relative to $z = 0$). The vertical axis corresponds to the angle between the host filament and the spin of this halo, and the horizontal axis to the redshift. Here, the spin becomes perpendicular to its filament after an important merger at redshift ~ 0.5. Right-hand panel: merging tree of a low-mass halo ($2 \times 10^{11}\,\mathrm{M}_\odot$ at redshift zero). Here, the spin becomes suddenly parallel to its host filament when the halo acquires most of its mass by accretion between redshift 1 and 2.

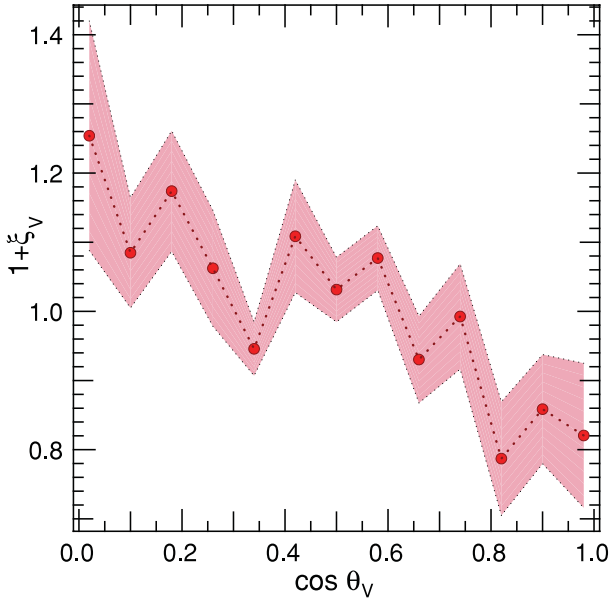


Figure 7. Excess probability of the velocity–filament alignment just before the formation time for a sample of 15 low-mass haloes ($M < M_{\rm crit}$). The average orientation over these 15 haloes is plotted with a dotted line, while the 1σ deviation away from that mean is displayed in pink. The progenitors' velocities before formation lie preferentially in a plane perpendicular to the closest filament (see also Figs 6 and A1).

less massive objects are born by accretion at high redshift during the winding of the walls into filaments. This process generates a spin aligned with the filaments. Most of the haloes of low mass at $z = 0$ are now formed, and their spin will not change much because they have already acquired most of their mass while future accretion will not be important enough to have a strong impact on their spin direction. This behaviour is illustrated in Fig. 6 (right-hand panel) where a halo forms and acquires a spin parallel to the direction of the forming filament, which then remains in the same direction during accretion until redshift zero.

At lower redshift, filaments are collapsing and thus create a flow along their direction in which the more massive haloes form by major mergers. During this process, these massive haloes acquire a spin which is the superposition of the spin of their progenitors and the orbital spin from the merger. As the motion is along the filament, this orbital spin is in a plane perpendicular to it. This process is shown in Fig. 6 (left-hand panel): the spin of the (less massive) progenitors is parallel to the filaments and their merger within the filaments induces a more massive halo whose spin is now perpendicular to the host filament. For these more massive DM haloes, a competition between orbital spin and intrinsic spin during the merger process has just been highlighted and explains the resulting orientation. Indeed, their spin is not randomly distributed in a plane perpendicular to the filaments but is shown to be correlated with one particular eigendirection of the large-scale tidal tensor. This issue is fully addressed in Appendix A.

It is interesting to compare the picture described above with Peirani, Mohayaee & de Freitas Pacheco (2004) who focused on the spin magnitude (instead of its direction). These authors claimed that spin acquisition was dominated by merger events, rather smooth accretion; Fig. 6 suggests that this holds true for the spin's direction of haloes more massive than the critical mass $M^{\rm s}_{\rm crit}$ as well.

3.3 Visual inspection using hydrodynamics

For illustrative purposes, let us now turn to a high-redshift hydrodynamical simulation, as the cold gas that we will analyse here follows more closely the caustics of the cosmic web than the DM, and thus provides a clearer visual impression of the process of wall winding and spin acquisition (in contrast to collisionless DM, cold gas does not undergo shell crossing but shocks and loses degrees of freedom – its motion perpendicular to the shock). We are not concerned here with how much respective angular momentum gas has with respect to DM acquired, but only use hydrodynamics as a proxy for pinpointing more accurately the loci of shell crossing and identifying the spin axis of galaxies. As we will argue later, it is also of interest to consider in parallel the environment of low-mass, high-redshift and high-mass, low-redshift haloes. Following Dubois et al. (2012) we use tracer particles of the gas in a cosmological hydrodynamical simulation (which is described in Appendix B) to illustrate this winding of walls and the loci and orientations of galaxies.

Fig. 8 represents the web-like filamentary structure of the gas at $z = 9$ in a field of size approximately $50\,h^{-1}$ kpc across; the ensemble of tracer particles initially makes up a sheet-like structure with a dominant filament embedded in it. Note that these tracer particles represent a biased subset of all tracers as they are chosen so as to end up within the bulge of the main galaxy of this zoom-in simulation at some later stage. The tracer particles flow from this sheet into the filaments where they form 'protogalaxies'. The gas thereby typically has a non-zero impact parameter relative to the filament and protogalaxies thus acquire a spin parallel to the filament in which they form (see the arrows in the figure). These young galaxies then migrate along their filaments and merge with other galaxies. The spin of the merger remnant is a combination of the orbital angular momentum of the collision and the initial spin of the progenitors; it can therefore depart from the direction of the host filament. We provide an animation at http://www.iap.fr/users/pichon/spin/ which allows this to be seen more easily (see also the Supporting Information with the online version of the paper). Fig. B1 illustrates quantitatively the visual impression of Fig. 8 while measuring the spin of the circumgalactic discs (between 0.1 and $0.25R_{\rm vir}$).

In Fig. 9, a small subset of these tracer particles are randomly chosen and followed for a while from early (left-hand panel) to late times (right-hand panel). On large scales (at early times), we note that the flow is indeed dominated by the winding up of matter from the main wall around the main filament. The tracers' trajectories start perpendicular to the filament within the walls. As they reach the filament, they take a sharp turn, losing their transverse motion and flowing along the filament (left-hand panel). In doing so, since the laminar flows on opposite sides of the wall will typically have different impact parameters, they generate a spinning structure whose axis will be aligned with the filament. This spaghetti structure converges into a quite narrow and elongated plait on either side of the forming disc. Given its orientation, the induced disc will therefore advect secondary infall at its periphery preferentially along its spin axis (as both the galaxy and its upcoming secondary infall were assembled via the same winding process).

In a nutshell, the cold gas dynamics of large-scale cosmic flows provide a much clearer illustration of the scenario we outlined for the formation of haloes with a spin mostly aligned with the local filament. We do not discuss any further the properties of the alignment of the disc with respect to the filamentary structure, given that there exist many caveats (though see Section 4.2 below and Appendix B).

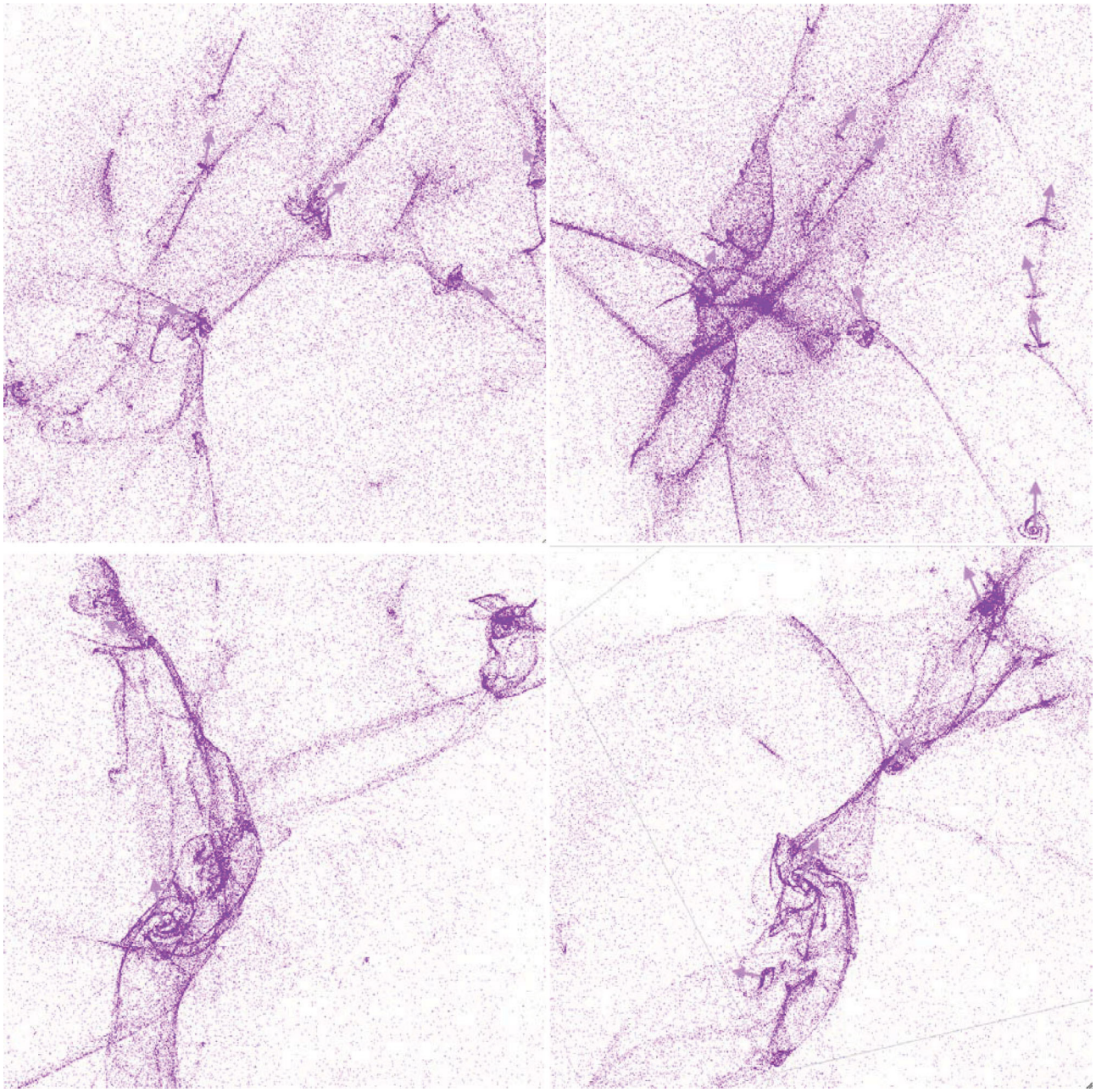


Figure 8. Distribution of hydrodynamical tracer particles (at $z \simeq 9$, top panels, and $z \simeq 8$, bottom panels). The web-like filamentary structure (top left-hand and right-hand panels) of the gas distribution which ends up in the bulge of a unique galaxy at later times is quite intricate, though one main wall in which the largest filament is embedded dominates (seen more clearly edge-on, in the top right-hand panel). Note the disc-like features with a spin parallel to the filament (represented qualitatively as an arrow perpendicular to the disc). Bottom panels: zoom-in at a later stage to visualize the process of a merger along the filament, before (bottom left-hand panel) and during (bottom right-hand panel) the merger. The spin of the merger remnant is a combination of the orbital angular momentum and the initial spin of the progenitors; it can therefore depart from the direction of the host filament. Note also the ribbon structure of filaments which corresponds to the locus of the second shock. Visual inspection suggests these ribbons become broader with time (as predicted by Pichon et al. 2011 as they advect larger and larger amount of angular momenta from the outskirts of the gravitational patch) and tend to lie perpendicular to the main wall; as they reach the protogalaxy, they twist rapidly on outer shells and build up its outskirts.

4 CONCLUSIONS AND PROSPECTS

In this paper, the Horizon 4π N-body simulation was used to investigate the correlations between the spin of DM haloes and their large-scale environment. For filaments defined over a smoothing scale of $5\,h^{-1}$ Mpc, a statistically significant signal was detected, indicating that the orientation of the spin of DM haloes is sensitive to the cosmic environment. A mass dependence of this signal was also robustly established: low-mass haloes are more likely to be aligned with large-scale filaments (with an excess probability of 15 per cent), whereas more massive haloes tend to be perpendicular to these (with an excess probability of 12 per cent). The mass transition was found to be redshift-dependent and to vary like $M_{\rm crit}(z) \approx M_0(1+z)^{-\gamma}$ with $M_0^{\rm s} \simeq 5(\pm 1) \times 10^{12}\,{\rm M}_\odot$ and $\gamma_{\rm s} = 2.5 \pm 0.2$. This critical mass is also found to increase with smoothing length (Appendix D1). These results are in agreement with those presented

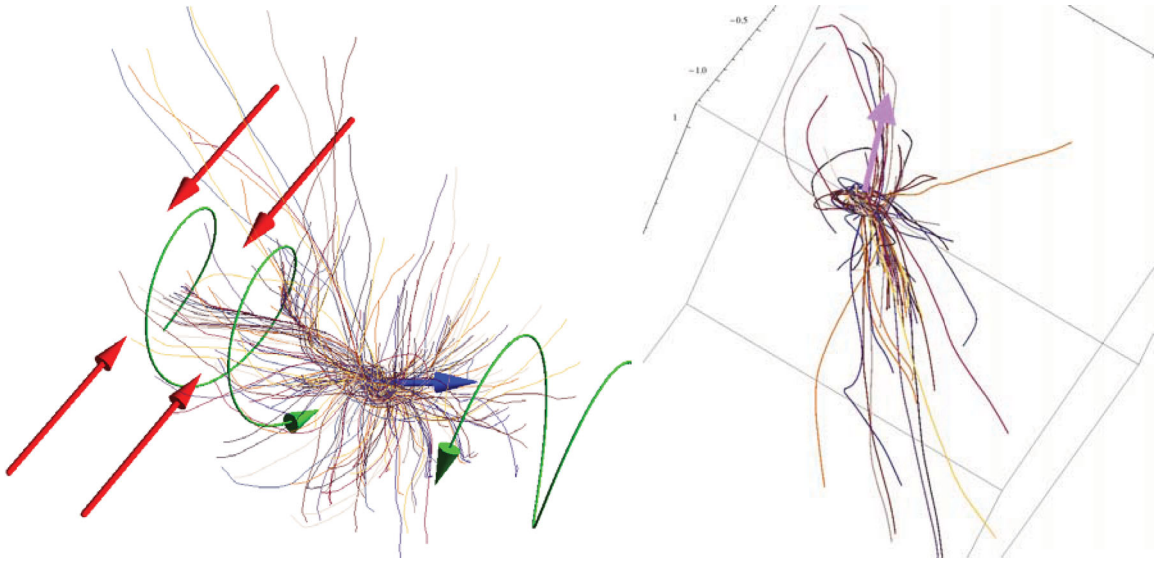


Figure 9. Trajectories of tracer particles in the outer region (left-hand panel) and inner region (right-hand panel). On large scales, the gas departs from voids, flows along the walls (red arrows) and winds up in filaments along well-defined ribbons (green arrows), forming low-mass haloes. In the inner region of the tracks, the flow is indeed along the filament and spiralling in into a disc whose axis is roughly aligned with the filament.

in Appendix A, which are derived using a more classic approach of considering excess alignment of the halo spin with the tidal tensor eigendirections. Since the tidal tensor probes larger scales than the Skeleton of the density, the mass transition is detected for a larger mass $M_0^{\rm t} \simeq 8(\pm 2) \times 10^{12}\,{\rm M}_\odot$, which scales slightly differently with redshift, that is, $\gamma_{\rm t} = 3 \pm 0.3$. Both redshift evolutions are roughly consistent with that of the formal non-linear mass scale.

4.1 Discussion

This unambiguous result confirms and quantifies recent findings (Bailin & Steinmetz 2005; Aragón-Calvo et al. 2007; Hahn et al. 2007b; Paz et al. 2008). It is also consistent with Sousbie et al. (2009) who found that the spin of DM haloes tends to be perpendicular to filaments. Indeed, weighting our statistics by the spin magnitude (which corresponds to their strategy), we recover their results. Zhang et al. (2009) found the same result, but did not have enough statistics to probe the signal above the critical mass. The study of Faltenbacher et al. (2002) focused on haloes above $10^{14}\,h^{-1}\,{\rm M}_\odot$, that is, above the critical mass at redshift zero, which is why, in agreement with this work, they found a trend for these haloes to have a spin perpendicular to their closest filament. In contrast, Hatton & Ninin (2001) considered haloes with mass around 10^{11}–$10^{12}\,{\rm M}_\odot$, that is, below the critical mass and thus found an alignment of the spins with the filaments, again in agreement with this work. This paper also confirms the very recent findings of Libeskind et al. (2012) who claimed that low-mass haloes (around 10^{10}–$10^{11}\,{\rm M}_\odot$) tend to be aligned with the filaments and to lie in the plane of walls.

Note that Appendix A predicts a strong trend for the spin vector of DM haloes to lie in the plane of large-scale walls. Such a signal was claimed to have been detected in observations by Lee & Erdogdu (2007), Trujillo et al. (2006), Navarro et al. (2004) and should be re-investigated in view of this paper's predictions. Furthermore, observers should now also be able to investigate the spin–filament correlations in a way directly comparable to the theoretical predictions for DM presented in this paper (Fig. 3). Indeed, the code DisPerSE (Sousbie 2011), which identifies filaments using persistence, can accurately deal with discrete and sparse data sets and should provide a good estimator for the direction of *observed* filaments.

The time evolution of the angular momentum of individual DM haloes obtained by following their progenitors using merger trees suggests that the spin direction results, on the one hand, from the winding of the walls into filaments (first-generation, low-mass haloes) and, on the other hand, from significant mergers occurring along those filaments (second-generation, more massive haloes). More specifically, the arguments we developed throughout this paper strongly suggest that the measured correlations can be understood as a consequence of the dynamics of large-scale cosmic flows. Indeed, low-mass haloes mostly form at high redshift within the filaments generated by colliding/collapsing walls. Such a process naturally produces a net halo spin parallel to the filaments. In contrast, high-mass haloes mainly form by merging with other haloes along the filaments at a later time when the filaments are themselves colliding/collapsing. Therefore, they acquire a spin which is preferentially perpendicular to these filaments. Visual examination of 'smoothed' DM and hydrodynamical simulations lends extra support to this picture (see also Appendix B). Measurements of the orientation of the spin relative to the eigenvectors of the tidal tensor are also consistent with such a scenario, provided one takes into account the fact that they probe typically larger scales of the density field.

From the point of view of a large-scale filament, most low-mass haloes are formed early from patches that are part of a planar, flattened inflow of matter on to that filament. For Gaussian random fields, the tidal tensor in such patches is correlated with the filament's direction (via the shape parameter γ, Pogosyan et al. 2009), resulting in the preferential alignment of the spin of such haloes along that filament. This process is related to the theoretical predictions of Pichon & Bernardeau (1999) who demonstrated, using the Zel'dovich approximation, how vorticity was generated during the first shell crossing. This vorticity will lie in the plane of the forming walls. Extending their predictions while focusing now on a 2D flow, we speculate that secondary shell crossing will lead to the formation

of vortices aligned with the forming filament (see figs 5 and 7 of Pichon & Bernardeau 1999, a possible section perpendicular to the axis of the filament).[2] In turn, these vortices could account for the spin of protogalaxies, as was suggested by the referee of that paper.

The excess probability in Figs 3 and A2 lies at the 15–40 per cent level. As such it mainly reflects a residual trend of coherence inherited from the large-scale cosmic environment in which haloes form. This does not preclude the multiscale hierarchical clustering process from erasing part of this more orderly dynamics. For instance, clustering and merging on smaller scales will in part perturb large-scale ordered ribbons (as is already visible in the bottom panels of Fig. 8).

The trend in Fig. 3 is found, on the one hand, in these measurements at the high end of the mass function at redshift zero (whose dynamics is only mildly non-linear) and, on the other hand, for the gas (a proxy for DM caustics) at very high z for lower mass galaxies (again not very far from linear dynamics at this epoch). Hence, we found in both regimes that some imprint of the LSS geometry and dynamics is directly responsible for the spin of the forming object and its post-merger transition. In this paper, we tried and explained the origin of the statistical signal, but we do not argue that large-scale dynamics dominate the non-linear regime of galaxy evolution on an individual object basis.

The remaining task involves understanding in detail the redshift dependence of the transition masses $M_{\rm crit}$ described by equations (1) and (2) [and also its dependence on the smoothing length, which might allow us to identify a scale at which the (anti-)alignment is strongest]. It would also clearly be of interest to further quantify (using larger samples) the findings of Hahn et al. (2010) regarding the alignment of the stellar disc/circumgalactic medium with respect to the LSS and investigate how these results depend on subgrid physics and feedback processes.

Let us conclude this paper with some speculations about what these results imply specifically for the process of galaxy formation at high redshift.

4.2 Implications for galaxy formation

We are now in a position to *speculate* about the implication of our findings for high-redshift galaxy formation. In the view of the robust measurements of Section 2 and Appendix A and the visual inspections of Section 3, it appears that (i) galaxies form preferentially along filaments; and (ii) their internal dynamics (hence their morphology) inherit important features from this anisotropic environment. The first point is backed by the distribution of filaments at the virial radius (appendix A of Pichon et al. 2011; Danovich et al. 2012, and indirectly by Fig. 1) and the second point by the measurements reported in this paper.

Indeed, filaments can be thought of as loci of constructive interferences from the long-wavelength modes of the initial power spectrum. On top of these modes, constructive interferences of high-frequency modes produce peaks which thus get a boost in density that allows them to pass the critical threshold necessary to decouple from the overall expansion of the Universe, as envisioned in the spherical collapse model (Gunn & Gott 1972). This well-known biased clustering effect has been invoked to justify the clustering of galaxies around the nodes of the cosmic web (White, Tully & Davis 1988). It also explains why galaxies form in filaments. In walls alone, the actual density boost is not sufficiently large to trigger galaxy formation. We therefore argue here that, statistically, the main nodes of the cosmic web are where galaxies migrate, not where they form; galaxies generally form while reaching filaments from walls. They thus inherit the anisotropy of their birth place as spin orientation. During migration, they may collide with other galaxies/haloes and erase part of their birth heritage when converting orbital momentum into spin via merger.

Recently, Pichon et al. (2011) showed how the cold gas drains out of the prominent voids in the cosmic web into sheets and filaments before it finally gets accreted on to DM haloes. Interestingly, the imprint of the larger scale pancake structure of the typical cosmic web around a filament and a peak allows us to be more specific about the geometry of this process. Indeed, one of the striking features of Figs 8 and 9 (probably best seen in the animation available as Supporting Information with the online version of the paper) is those ribbon-like caustics which feed the central galaxy along its spin axis from both poles. Generically, the gas inflow in the frame of the galaxy is double-helix like along its spin axis; this is mostly wiped out in DM (and hardly visible in Fig. 1) because of shell crossing, but quite visible for the gas. These ribbons are generated via the same winding/folding process as for the protogalaxy, and represent the dominant source of secondary filamentary infall described in Pichon et al. (2011), which feeds the newly formed galaxy with gas of well-aligned angular momentum (whose direction was set by the impact parameter offset of the two neighbouring walls, which can in turn be attributed to the dissymmetry of the four neighbouring voids). As such, the larger scale geometry of the LSS (which biases the formation process) squashes the average neighbourhood of a peak (six saddles, eight voids, twelve walls) into a simpler effective geometry (one wall and one embedded filament dominating). Formally, the most likely 'crystal' of the Universe – subject to the constraint of collapse along two axes on larger scales – differs from the azimuthally averaged cubic centred crystal found in Pichon et al. (in preparation): it is quite flattened and dominated by one ridge (Pichon et al. 2011; Danovich et al. 2012; Dubois et al. 2012). Note that the gas flowing roughly parallel to the spin axis of the disc along both directions will typically impact the disc's circumgalactic medium and shock once more (as it did when it first reached the wall, and then the filaments, forming those above-mentioned ribbons), radiating away its vertical momentum (see Fig. 10 and Tillson, Miller & Devriendt 2011).

Our speculations here have focused on a two-scale process. Given the characteristics of ΛCDM hierarchical clustering, one can anticipate that this process occurs on several nested scales at various epochs – and arguably on various scales at the same epoch.[3] In other words, one expects smaller scale filaments are themselves embedded in larger scale walls (as discussed in Appendix A to reconcile our excess alignment with the eigenvectors of the tidal tensor). The induced multiscale anisotropic flow transpires in the scaling of transition mass with the smoothing length, as discussed in Appendix D1.

Another issue would be to estimate for how long this entanglement between the large-scale dynamics and the kinematic properties of high redshift pervades. Indeed, Ocvirk, Pichon & Teyssier (2008)

[2] Note, however, that such a typical caustic should have inherited some level of asymmetry (Pichon et al. 2011), which could imply that one vortex dominates.

[3] The scenario we propose for the origin of this signal is, like the signal itself, relative to the linear scale involved in defining the filaments and, as such, multiscale. It will hold as long as filaments are well defined in order to drive the local cosmic flow.

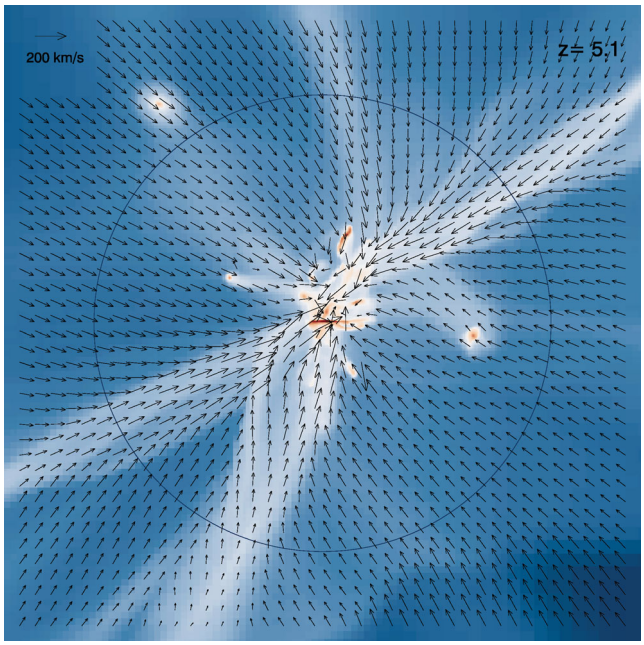


Figure 10. A meridional projection through the velocity flow and gas density around the NUT galaxy at redshift 5.1 (Kimm et al. 2011). The disc plane is along the horizontal axis and the circle marks the virial radius (17.6 kpc). Note the ribbon-like cold flows seen directly in the gas density and the velocity flow, which hit the galactic disc roughly along its spin axis. Note also how the gas takes a sharp turn when it reaches the ribbons.

have shown that at lower redshift the so-called hot mode of accretion will kick in; how will hot flows wash out/disintegrate these ribbons? Given that they locally reflect the large-scale geometry, will the gas continue to flow in along preferred directions (as DM does, see e.g. Aubert, Pichon & Colombi 2004), or does the hot phase erase any anisotropy? Will the above-mentioned smaller scale non-linear dynamics eventually wash out any such trace?

Finally, note that the actual spin of the stellar disc at low redshift need not be trivially related to that of its larger scale gravitational patch (see, for instance, Hahn et al. 2010), as a significant amount of angular momentum redistribution takes place in the circumgalactic medium (Kimm et al. 2011) over cosmic time.

ACKNOWLEDGMENTS

We thank J. Binney, M. Haehnelt, S. Peirani, S. Prunet and T. Kimm for advice, and our collaborators of the Horizon Project (www.projet-horizon.fr) for helping us produce the Horizon 4π simulation. We also thank the anonymous referee for constructive criticism. The hydrodynamical simulation presented here was run on the DiRAC facility jointly funded by STFC, the Large Facilities Capital Fund of BIS and the University of Oxford. CP acknowledges support from a Leverhulme visiting professorship at the Department of Physics of the University of Oxford, and thanks the Merton College, Oxford, for a visiting fellowship. DP thanks the French Canada Research Fund and the University of Oxford. JD and AS's research is supported by Adrian Beecroft, the Oxford Martin School and STFC. Special thanks to T. Kimm for Fig. 10, F. Bouchet for allowing us to use the Magique 3 supercomputer during commissioning, and to S. Rouberol for making it possible. We also thank D. Munro for freely distributing his YORICK programming language and OPENGL interface (available at http://yorick.sourceforge.net/).

REFERENCES

Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007, ApJ, 655, L5
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bailin J., Steinmetz M., 2005, ApJ, 627, 647
Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, MNRAS, 216, 17
Bond J. R., Myers S. T., 1996, ApJS, 103, 1
Bond J. R., Kofman L., Pogosyan D., 1996, Nat, 380, 603
Bond N. A., Strauss M. A., Cen R., 2010, MNRAS, 409, 156
Catelan P., Theuns T., 1996, MNRAS, 282, 436
Crittenden R., Natarajan P., Pen U., Theuns T., 2001, ApJ, 559, 552
Danovich M., Dekel A., Hahn O., Teyssier R., 2012, MNRAS, 422, 1732
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dekel A., 1985, ApJ, 298, 461
Desjacques V., 2008, MNRAS, 388, 638
Doroshkevich A. G., 1970, Astrofizika, 6, 581
Dubois Y., Teyssier R., 2008, A&A, 477, 79
Dubois Y., Pichon C., Haehnelt M., Kimm T., Slyz A., Devriendt J., Pogosyan D., 2012, MNRAS, 423, 3616
Faltenbacher A., Gottlöber S., Kerscher M., Müller V., 2002, A&A, 395, 1
Flin P., Godlowski W., 1986, MNRAS, 222, 525
Flin P., Godlowski W., 1990, Sov. Astron. Lett., 16, 209
Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, MNRAS, 396, 1815
Godłowski W., Flin P., 2010, ApJ, 708, 920
González R. E., Padilla N. D., 2010, MNRAS, 407, 1449
Gunn J. E., Gott J. R., III, 1972, ApJ, 176, 1
Haardt F., Madau P., 1996, ApJ, 461, 20
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007a, MNRAS, 375, 489
Hahn O., Carollo C. M., Porciani C., Dekel A., 2007b, MNRAS, 381, 41
Hahn O., Teyssier R., Carollo C. M., 2010, MNRAS, 405, 274
Hatton S., Ninin S., 2001, MNRAS, 322, 576
Hoyle F., 1949, in Burgers J. M., van de Hulst H. C., eds, Problems of Cosmical Aerodynamics. Central Air Documents Office, Dayton, OH, p. 195
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Icke V., 1973, A&A, 27, 1
Kimm T., Devriendt J., Slyz A., Pichon C., Kassin S. A., Dubois Y., 2011, preprint (arXiv:e-prints)
Komatsu E. et al., 2011, ApJS, 192, 18
Lee J., Erdogdu P., 2007, ApJ, 671, 1248
Lee J., Pen U., 2000, ApJ, 532, L5
Lee J., Pen U.-L., 2002, ApJ, 567, L111
Lemson G., 1993, MNRAS, 263, 913
Libeskind N. I., Hoffman Y., Knebe A., Steinmetz M., Gottlöber S., Metuki O., Yepes G., 2012, MNRAS, 421, L137
Lynden-Bell D., 1964, ApJ, 139, 1195
Navarro J. F., Abadi M. G., Steinmetz M., 2004, ApJ, 613, L41
Novikov D., Colombi S., Doré O., 2006, MNRAS, 366
Ocvirk P., Pichon C., Teyssier R., 2008, MNRAS, 390, 1326
Paz D. J., Stasyszyn F., Padilla N. D., 2008, MNRAS, 389, 1127P
Peebles P. J. E., 1969, ApJ, 543, L107
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Peirani S., Mohayaee R., de Freitas Pacheco J. A., 2004, MNRAS, 348, 921
Pichon C., Bernardeau F., 1999, A&A, 343, 663
Pichon C., Pogosyan D., Kimm T., Slyz A., Devriendt J., Dubois Y., 2011, MNRAS, 418, 2493
Platen E., van de Weygaert R., Jones B. J. T., 2007, MNRAS, 380, 551
Platen E., van de Weygaert R., Jones B. J. T., 2008, MNRAS, 387, 128
Pogosyan D., Bond J. R., Kofman L., Wadsley J., 1998, in Colombi S., Mellier Y., Raban B., eds, Wide Field Surveys in Cosmology Cosmic Web: Origin and Observables. Editions Frontières, Gif-sur-Yvette, p. 61
Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F., Sousbie T., Colombi S., 2009, MNRAS, 396, 635
Porciani C., Dekel A., Hoffman Y., 2002, MNRAS, 332, 325

Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R., Gottloeber S., 2008, ApJS, 178, 179
Rasera Y., Teyssier R., 2006, A&A, 445, 1
Schäfer B. M., 2009, Int. J. Mod. Phys. D, 18, 173
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Slosar A., White M., 2009, J. Cosmol. Astropart. Phys., 6, 9
Sousbie T., 2011, MNRAS, 414, 350
Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008, MNRAS, 383, 1655
Sousbie T., Colombi S., Pichon C., 2009, MNRAS, 393, 457
Spergel D. N. et al., 2003, ApJS, 148, 175
Stoica R. S., Martìnez V. J., Mateu J., Saar E., 2005, A&A, 434, 423
Sugerman B., Summers F. J., Kamionkowski M., 2000, MNRAS, 311, 762
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Teyssier R., 2002, A&A, 385, 337
Teyssier R. et al., 2009, A&A, 497, 335
Tillson H., Miller L., Devriendt J., 2011, MNRAS, 417, 666
Trujillo I., Carretero C., Patiri S. G., 2006, ApJ, 640, L111
Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647
Varela J., Betancort-Rijo J., Trujillo I., Ricciardelli E., 2012, ApJ, 744, 82
White S. D. M., 1984, ApJ, 286, 38
White S. D. M., Silk J., 1979, ApJ, 231, 1
White S. D. M., Tully R. B., Davis M., 1988, ApJ, 333, L45
Zel'Dovich Y. B., 1970, A&A, 5, 84
Zeldovich I. B., Einasto J., Shandarin S. F., 1982, Nat, 300, 407
Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747

APPENDIX A: SPIN–TIDAL TENSOR CORRELATIONS

Let us present a complementary set of measurements: the correlations between the spin axis of dark haloes and the orientation of the large-scale gravitational tidal tensor $T_{ij} = \partial_{ij}\phi - \frac{1}{3}\Delta\phi\delta_{ij}$. To describe the orientation of the tidal tensor, we define $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ to be the minor, intermediate and major eigendirections of T_{ij} according to the sorted eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ of the Hessian of the gravitational potential, $\partial_{ij}\phi$ (with which the tidal tensor shares the eigendirections).

Besides works based on the correlations between spin orientation and the cosmic web described in the main text (Section 1), the only numerical study to date looking at the alignment between halo spin and tidal tensor was done by Porciani, Dekel & Hoffman (2002) who predicted its orthogonality with the major principal axis but also found that galactic spins must have lost their initial alignment with the tidal tensor predicted by TTT. Direct observations of the alignment between the spin and the tidal tensor eigenvectors have also been carried out: the first attempt by Lee & Pen (2002) studied the correlations between the disc orientation of the galaxies from the Tully catalogue and the shear reconstructed from the Point Source Catalog Redshift Survey and rejected the hypothesis of randomness at a 99.98 per cent confidence level. More recently, Lee & Erdogdu (2007) detected some correlations between the spin and the intermediate eigenvector of the tidal tensor and found that galactic spins were also preferentially perpendicular to the major principal axis but this signal remains weak. To overcome this lack of a clear numerical detection, the use of the 43 million halo sample of the Horizon 4π simulation presents a tremendous advantage, as it allows us to very robustly calculate the correlations between halo spin orientation and the local tidal tensor. We show in this appendix that these measurements are not only consistent with the spin–filament correlations, but actually lend additional support to the interpretation of this paper in terms of large-scale dynamics along walls and filaments.

At the onset of non-linearity, the gravitational potential tracks the velocity potential of the matter flow. Thus, the signs of λ_i determine whether the flow in the corresponding direction compresses ($\lambda_i > 0$) or rarifies ($\lambda_i < 0$) matter. At a given smoothing scale, this criterion can be used to partition space into peak-like $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3$, filament-like $\lambda_1 < 0$, $0 < \lambda_2 \leq \lambda_3$, wall-like $\lambda_1 \leq \lambda_2 < 0$, $0 < \lambda_3$ or void-like $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq 0$ regions (Pogosyan et al. 1998). From this point of view, in the peak regions matter compression is strongest along $\boldsymbol{e}_3$ and weakest along $\boldsymbol{e}_1$. In the filamentary regions, $\boldsymbol{e}_1$ gives the direction of the filament, while the walls are collapsing along $\boldsymbol{e}_3$ and extend, locally, in the plane spanned by $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$.

At this stage, it is important to note that the tidal field probes larger scale structures than the filamentary structure studied in Section 3 as the gravitational potential is a smoother version of the density field (through Poisson's equation). In other words, the Skeleton of the potential (which locally corresponds to the eigendirections of the tidal tensor) traces the cosmic structures (walls, filaments, etc.) on scales much larger than the Skeleton of the density. Thus, if we turn to a formulation in terms of the Skeleton of the potential, the filaments described in Section 3 are embedded in the large-scale walls of the potential field (as illustrated in Fig. A1); therefore, protogalaxy formation begins with the first collapse (namely the collapse of $\boldsymbol{e}_3$) leading to the formation of the large-scale walls because it corresponds on smaller scale to the time when the filaments (of the density field) form by winding.

Note that a stricter definition of a filament in the local theory of the Skeleton (Sousbie et al. 2008) as a ridge in the density profile associates its local direction with the minor eigendirection of the Hessian of the *density* (here, to match enumeration, taken with a negative sign) $\boldsymbol{e}_1^{\rho}$. Thus, as the potential is two derivatives away from the density, the excess probability of alignment between the

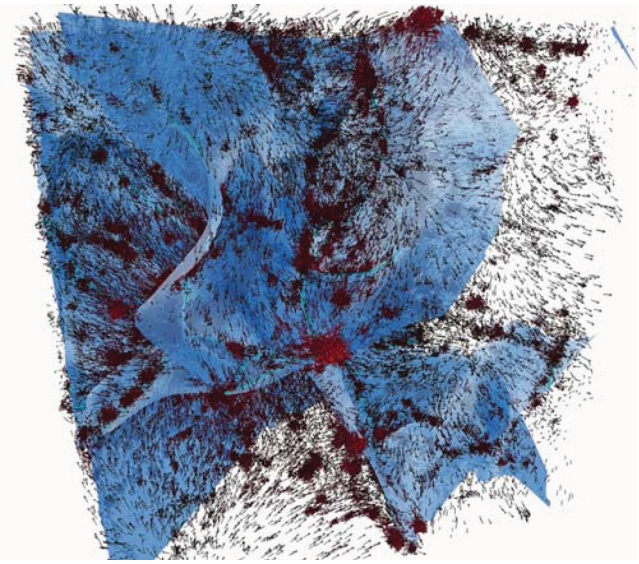


Figure A1. The walls of the potential (in blue) and its filaments (in cyan), together with the velocities of DM particles (in red) of a $20\,h^{-1}$ Mpc 512^3 DM particle ΛCDM simulation with *WMAP1* cosmogony. At redshift zero, most DM haloes sit in these walls, while the velocity field empties the voids and flows within those walls. This divergent flow is best seen in the top (hence bottom) left-hand void (see also the animations available as Supporting Information with the online version of the paper).

halo spin and $\boldsymbol{e}_1$ should be quite similar to that of the alignment between the halo spin and the filament's direction.

A1 Alignment between spin and tidal eigendirections

In order to compute the excess probability of alignment between the spin of DM haloes and the eigenvectors of the tidal tensor, the density field is again smoothed over $5\,h^{-1}$ Mpc (for $z = 0$), Poisson's equation is solved via fast Fourier transform in smaller overlapping boxes (boundary effects are found to be insignificant inside the boxes), the Hessian matrix of the potential is computed using a finite difference scheme and this matrix is finally interpolated to halo positions. We then measure the angle between the angular momentum vector of the halo and each eigendirection of the tidal tensor and compute the histogram of the absolute value of the cosine of this angle; after normalization, it gives $1 + \xi$, the excess probability of alignment between the spin and the tidal tensor eigendirections. As in Section 2.2, the data are then split by halo mass.

Fig. A2 displays these excess probabilities in three panels corresponding to the orientation of the spin with respect to $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ (top, middle and bottom panels, respectively), at $z = 0$. Haloes of all masses have spins that preferentially avoid the direction of the strongest large-scale compression, $\boldsymbol{e}_3$, at 10 per cent excess probability (bottom panel). The exception is the very highest mass bin, which shows an additional halo population with spins aligned with $\boldsymbol{e}_3$.

For the high-mass, $M > M_0^{\rm t} \approx 8(\pm2) \times 10^{12}\,{\rm M}_\odot$, haloes we detect a strong trend for the spin to be aligned with $\boldsymbol{e}_2$, the intermediate principal axis of the tidal tensor with an excess probability of up to 40 per cent (blue and green lines, middle panel), and to be perpendicular to the minor, $\boldsymbol{e}_1$, principal axis with an excess probability of up to 20 per cent (blue and green lines, top panel), a result in agreement with Lee & Erdogdu (2007) and Porciani et al. (2002).

Spins of the lower mass, $M < M_0^{\rm t}$, haloes tend, in contrast, to align with $\boldsymbol{e}_1$ (red and orange lines, top panel), thus preferring the direction of the filamentary structures $\boldsymbol{e}_1$, with an excess probability of up to 15 per cent, and in a weaker way to align with $\boldsymbol{e}_2$ (with an excess probability of 5 per cent, see the red and orange lines in the middle panel).

These results are in exact agreement with our findings using the Skeleton that the spins of sufficiently large haloes prefer to be perpendicular to the filament's direction, while small haloes show a positive correlation for a spin orientation along the filaments. Indeed, the bottom panel of Fig. A2 is almost identical to the spin–filament correlation found in Fig. 3. For the tidal tensor, the transition occurs at a somewhat higher mass, $M_0^{\rm t} \approx 8 \times 10^{12}\,{\rm M}_\odot$, than for the Skeleton probe. This is not surprising because of the effectively larger scales probed by the tidal tensor and the observation (see Section 2.2 and Appendix D1) that the critical mass increases with the smoothing length used to define the LSS.

The redshift dependence of the transition mass was also investigated (as described in the main text, see Section 2.3) and is found to be

$$M_{\rm crit}^{\rm t} \approx M_0^{\rm t}(1+z)^{-\gamma_{\rm t}}, \quad \gamma_{\rm t} = 3 \pm 0.3, \quad M_0^{\rm t} \simeq 8(\pm2) \times 10^{12}\,{\rm M}_\odot.$$

A2 Consistency with LSS cosmic flows

Let us describe this spin acquisition in the framework of large-scale dynamics and the ellipsoidal collapse model (Lynden-Bell 1964; Zel'Dovich 1970; Icke 1973; White & Silk 1979; Peebles 1980;

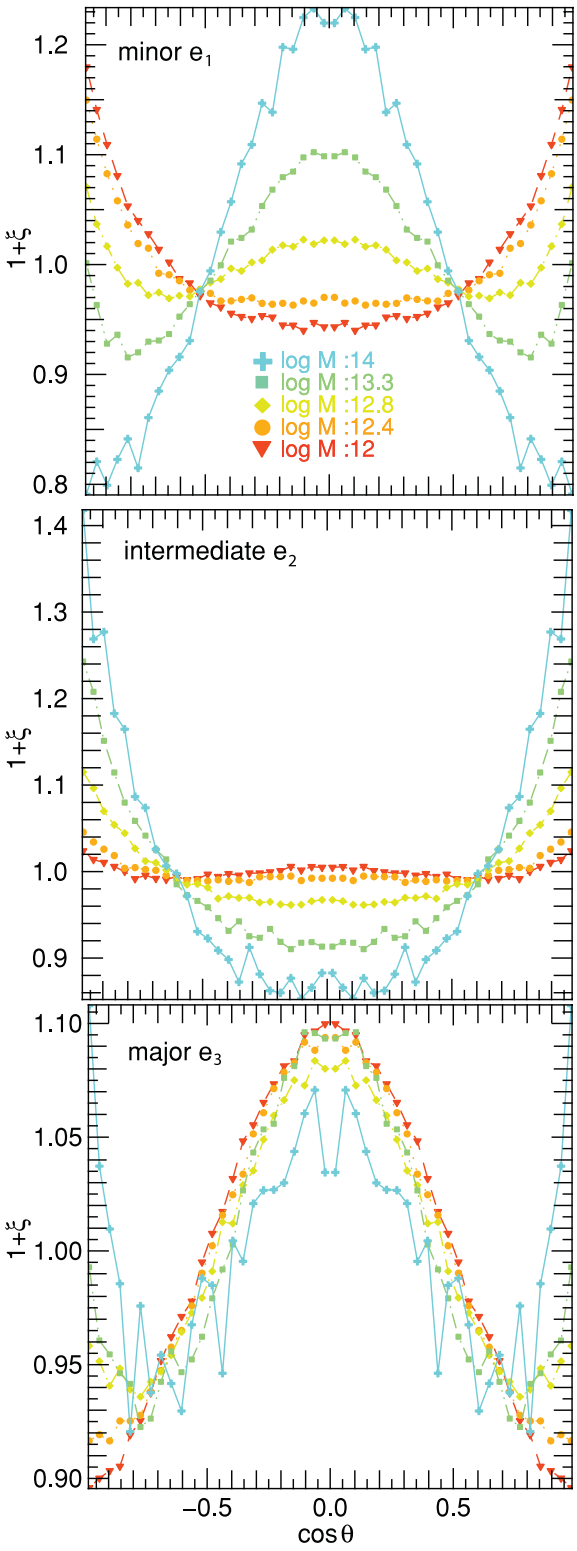


Figure A2. Excess probability of alignment between the spin and the minor/intermediate/major axis (from the top to bottom) of the tidal tensor in the Horizon 4π simulation. Different mass bins are colour-coded from 10^{12} (red) to $10^{14}\,{\rm M}_\odot$ (blue). A transition is detected: the spin of high-mass haloes tends to be aligned with the intermediate principal axis (middle panel), whereas the spin of low-mass haloes tends more likely to point along the minor axis (top panel).

Lemson 1993; Bond & Myers 1996; Sheth, Mo & Tormen 2001; Desjacques 2008, among others). Low-mass haloes form by accretion at high redshift. At this time, sheets and filaments are forming by the successive collapse of $\boldsymbol{e}_3$ and $\boldsymbol{e}_2$, respectively. When $\boldsymbol{e}_3$ is collapsing, low-mass haloes form by accreting particles whose motion is along the direction $\boldsymbol{e}_3$. This process induces a spin perpendicular to this direction, that is, in the plane $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ which is the plane of the large-scale wall in which they are located. Then, $\boldsymbol{e}_2$ begins to collapse and other low-mass haloes form from objects moving in the plane $(\boldsymbol{e}_2, \boldsymbol{e}_3)$, thus acquiring a spin perpendicular to this plane, that is, aligned with $\boldsymbol{e}_1$ (which is the direction of the forming large-scale filament). The first generation of DM haloes (of typically low mass) are now formed; as mentioned in Section 3, for low-mass haloes at redshift zero, the spin will not change much as they have already acquired most of their mass. Fig. 6 (left-hand panel) provides a clear illustration for this: a low-mass halo forms and acquires a spin aligned with $\boldsymbol{e}_1$ as expected. It has been pointed out above that halo spin orientation must be either aligned with $\boldsymbol{e}_1$ (correlated to the filament's direction), or in the plane $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ depending on the time when they form. This is in good agreement with the top and middle panels of Fig. A2 which show an excess probability for their spin to be aligned with $\boldsymbol{e}_1$ (or with the filaments) and in a weaker way with $\boldsymbol{e}_2$ (see the red and orange lines which represent the low-mass haloes at redshift zero, that is, the haloes with a mass between 3×10^{11} and $3 \times 10^{12}\,\mathrm{M}_\odot$).

Later (as described in Section 3 in terms of flows along the filaments), $\boldsymbol{e}_1$ collapses and the haloes located in the filaments of the potential stream along this direction. Most of the more massive haloes then form by mergers in this flow and therefore acquire a spin which combines the spin of their progenitors and the orbital spin provided by the merger. The orbital spin must be in the plane perpendicular to $\boldsymbol{e}_1$, that is, $(\boldsymbol{e}_2, \boldsymbol{e}_3)$ because the progenitors move along the filaments before merging, whereas the spin of their (less massive) progenitors is in the plane $(\boldsymbol{e}_1, \boldsymbol{e}_2)$. The resulting angular momentum is therefore a superposition of these various spins, statistically more likely to be aligned with $\boldsymbol{e}_2$, which is what shown in Fig. 6 (right-hand panel) and which is also in good agreement with the middle panel of Fig. A2, where the blue and green lines representing haloes above $2 \times 10^{13}\ \mathrm{M}_\odot$ reveal a strong trend for these high-mass haloes to be aligned with $\boldsymbol{e}_2$. For these massive DM haloes, the competition between the orbital spin and the intrinsic spin during the merger process was already pointed out in Section 3. The excess probability for their spin to be aligned with $\boldsymbol{e}_2$ suggests that neither one nor the other dominates. Nevertheless, very massive haloes (with masses above $10^{14}\,\mathrm{M}_\odot$) represented with a blue line in Fig. A2 seem to have their spin less perpendicular to $\boldsymbol{e}_3$: actually we can observe two modes, one perpendicular and one aligned with this direction, which can be understood if they are the result of a further generation of mergers whose intrinsic spins were already perpendicular to the filament.

Let us emphasize that this explanation and that which was presented in Section 3 are consistent. Indeed, in Section 3 our claim is that the winding of the walls is responsible for the direction of the spin of low-mass haloes. Meanwhile, in this appendix, the focus is on the first collapse along $\boldsymbol{e}_3$, that is, on the formation of walls. However, as pointed out in the introduction of this appendix, the tidal field probes *larger* scale structures than the filaments of the density field studied in Section 3. Taking into account this difference, it turns out that the two analyses are complementary (describing the large-scale dynamics on different scales). It also helps reconciling the findings of Porciani et al. (2002) and Lee & Erdogdu (2007) which rely on the tidal tensor with those of Bailin & Steinmetz (2005), Aragón-Calvo et al. (2007), Hahn et al. (2007b) and Paz et al. (2008) which involve the density field filamentary structure.

APPENDIX B: CIRCUMGALACTIC MEDIUM SPIN

Let us first describe briefly the *hydrodynamical* simulation that we will use to assess the alignment of the gas component surrounding the galaxies with the large-scale filamentary structure. This simulation is described in more details in Dubois et al. (2012) and corresponds to their SHhr run. Let us recall here its basic properties. The SHhr simulation follows the formation of a massive ($M_{\rm vir} = 5 \times 10^{11}\,\mathrm{M}_\odot$ at $z = 6$) halo within a resimulated region in a $100\,h^{-1}$ Mpc box size, its DM mass resolution is $1.3 \times 10^6\,\mathrm{M}_\odot$, and minimum cell size is 17 pc. The cosmology employed in that run is slightly different from the parameters of the Horizon 4π simulation, and is compatible with the *WMAP7* cosmology (Komatsu et al. 2011): $\Omega_{\rm m} = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_{\rm b} = 0.045$, $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $n = 0.961$ and $\sigma_8 = 0.8$. The gas is allowed to cool down radiatively down to $T_0 = 100\,\mathrm{K}$, assuming an initial metal enrichment of $10^{-3}\,\mathrm{Z}_\odot$ (Sutherland & Dopita 1993). An ultraviolet background heating source term is added to the gas energy equation following Haardt & Madau (1996) with reionization taking place at $z_{\rm reion} = 8.5$. Star formation is allowed in gas density regions above $n_0 = 50\,\mathrm{H\,cm^{-3}}$ using a Poisson random process (Rasera & Teyssier 2006; Dubois & Teyssier 2008) that reproduces the Schmidt–Kennicutt law $\dot{\rho}_* = \epsilon_* \rho / t_{\rm ff}$, where $\dot{\rho}_*$ is the star formation rate density, $\epsilon_* = 0.01$ is the star formation efficiency and $t_{\rm ff}$ is the local free-fall time of the gas with local density ρ. No feedback from supernovae or active galactic nuclei is accounted for. Tracer particles that passively follow the motion of the gas are scattered in the initial conditions and allow us to trace back the Lagrangian trajectories of gas elements that end up in collapsed structures.

Fig. B1 displays two large-scale views of this hydrodynamical simulation and its tracer particles at redshift 9 shown in Fig. 8. Structures and substructures in DM are detected with the Most massive Sub-node Method (Tweed et al. 2009) and only masses above $5 \times 10^8\,\mathrm{M}_\odot$ are selected. The spin of the circumgalactic medium (accounting for non-star-forming gas only with gas density below $n_{\rm H} < 50\,\mathrm{H\,cm^{-3}}$) between 0.1 and $0.25 R_{\rm vir}$ is then computed and its orientation is represented with the dark red segments. This somewhat ad hoc criterion used to define the spin reflects our focus on the angular momentum of the secondary infall gas, which has just been or is being accreted; our measurements correspond to an average of the spin within this subregion of the DM halo. This figure shows a good alignment of the spins with the circumgalactic polar filaments (i.e. the filament which visually flows along the polar axis of the galaxy, in particular for clumps 2, 3, 4 and 5). This is consistent with visual inspection (Fig. 8, but best viewed via the online Supporting Information) and with the prediction of Fig. 3, since the critical mass at that redshift is $6 \times 10^{10}\,\mathrm{M}_\odot$. The spin of a couple of low-mass clumps in that field (noticeably clumps 1 and 6) is in fact poorly estimated automatically, as the DM clump centre can be offset at that redshift relative to that of the circumgalactic disc.

APPENDIX C: ROBUSTNESS OF THE SPIN–FILAMENT CORRELATION

Let us assess how robust the spin orientation–filament correlation found in Section 2.2 is by carrying out a few consistency tests.

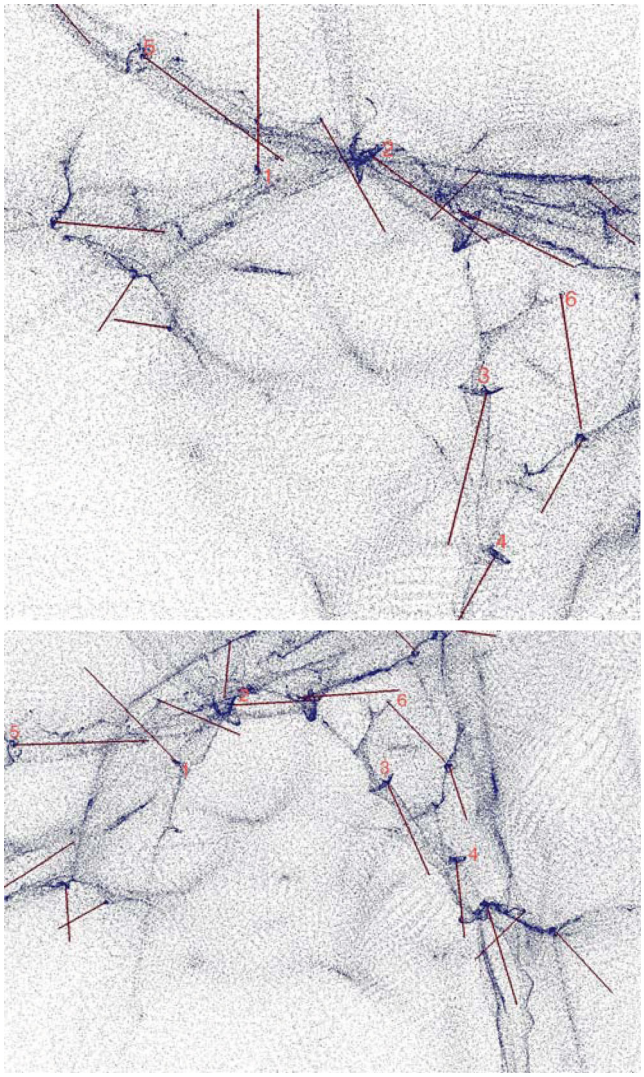


Figure B1. Distribution of hydrodynamical tracer particles (in blue) at redshift 9 along two different viewing angles. The spin direction of the circumgalactic medium between 0.1 and $0.25R_{\rm vir}$ of some protogalaxies is plotted in red. On average the spins are more likely to be aligned with the filaments (as seen with protogalaxies 2, 3, 4 and 5, for example).

In order to check the effect of mass resolution (the spin of very low mass haloes is poorly defined, for instance, as too few particles are involved in its measurement), the same measurements are carried out in a smaller simulation (256^3 particles in a $50\,h^{-1}$ Mpc periodic box with the same cosmology) for which same physical masses are represented by higher numbers of particles: mass bins from 3×10^{11} (red) to $6 \times 10^{12}\,{\rm M}_\odot$ (green) correspond to 30–1400 particles in Horizon 4π (corresponding to a lower threshold for the FoF detection) since the mass per particle is $7 \times 10^9\,{\rm M}_\odot$; in contrast, 350–15 000 particles are found in haloes of the same mass for the small simulation since the mass per particle is $6 \times 10^8\,{\rm M}_\odot$. The detection of the same phase transition (see Fig. C1) occurring at the same halo mass demonstrates that the signal is not induced by limited mass resolution. This result is also consistent with the findings of Aragón-Calvo et al. (2007) and Hahn et al. (2007b).

Another simple check involves varying the procedure by choosing for each halo the closest segments of the Skeleton (instead of for each segment, the closest haloes). The signal we get is very similar to Fig. 3 (low-mass haloes tend to be parallel to the filaments with

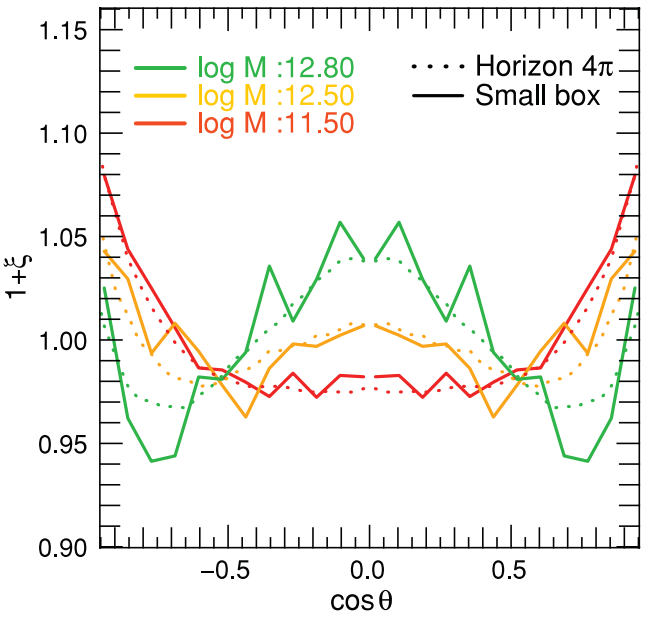


Figure C1. Excess probability of spin–filament alignment in various simulations. Measurements in the Horizon 4π simulation are plotted with the dotted lines and measurements in the smaller simulations are plotted with the solid lines. Different colours correspond to different halo mass bins from 3×10^{11} (red) to $6 \times 10^{12}\,{\rm M}_\odot$ (green). The signal in the Horizon 4π simulation is statistically consistent with that of the smaller simulations.

an excess probability of 15 per cent and high-mass haloes perpendicular with an excess probability reaching 20 per cent for the more massive bin which is even stronger than in Fig. 3), suggesting that the measured correlations are independent of the detailed procedure implemented to identify neighbours.

Haloes at the nodes of the Skeleton cannot have a well-defined closest segment direction, as more than one Skeleton segment typically qualifies, and could therefore bias our measurements. So as to quantify this effect, the same algorithm is implemented but with a new criterion: haloes closer than a certain distance to the nodes are not considered. An even stronger signal is detected, which leads us to conclude that nodes introduce extra noise and are not the cause for the observed signal.

One might also think this result could depend on a density threshold for the underlying filament. Hence, the same data in low-, intermediate- and high-density filaments were plotted: the excess probability of spin–filament alignment is found to be the same whatever the density inside the filaments is. All these tests demonstrate the overall robustness of the mass-dependent transition of the relative alignment.

APPENDIX D: CHARACTERISTIC MASSES

In the main text, the density field was smoothed over a scale of $5\,h^{-1}$ Mpc corresponding to a mass of $1.9 \times 10^{14}\,{\rm M}_\odot$. The transition mass found in Section 2.2 is therefore defined relatively to this mass. In the context of hierarchical clustering, as long as the field is smoothed on scales where filaments are still well defined, one can anticipate some scaling of this transition mass with smoothing. This transition mass should reflect the connection between the geometry of the larger scale flow and the mass scale corresponding to galaxies forming and drifting on this cosmic web.

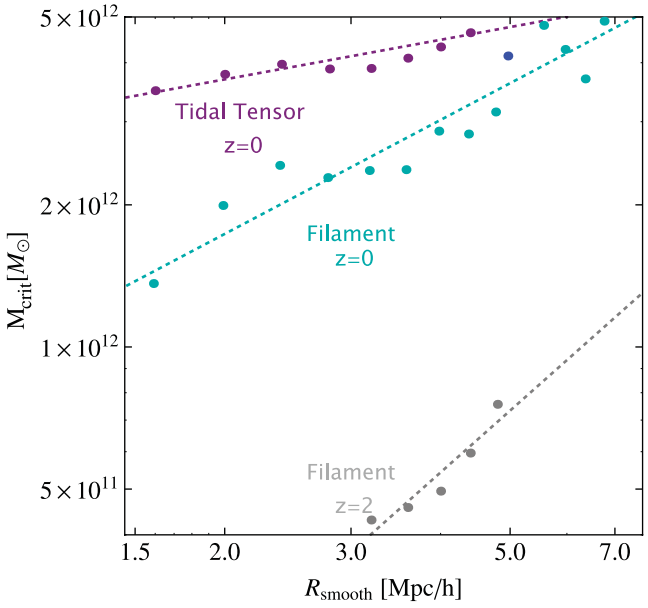


Figure D1. Evolution of the critical mass as a function of the smoothing length for the filaments at redshift zero (cyan) and two (grey) and the tidal tensor at redshift zero (purple). The critical mass increases with the smoothing length with a redshift-dependent slope. The blue dot was measured in the Horizon 4π simulation for the filaments.

D1 Smoothing dependence of the critical mass

Fig. D1 displays the evolution of the critical mass with smoothing length. It is found that $M^{\rm s}_{\rm crit}(R)$ and $M^{\rm t}_{\rm crit}(R)$ can be well fitted by power laws, namely

$$M_0(R) \simeq M_0(R_0)\left(\frac{R}{R_0}\right)^{\alpha}, \qquad \text{(D1)}$$

where $M^{\rm s}_0(R_0) \simeq 3.6(\pm1)\times10^{12}$, $\alpha^{\rm s} \simeq 0.8 \pm 0.1$, $M^{\rm t}_0(R_0) \simeq 4.8(\pm1)\times10^{12}$ and $\alpha^{\rm t} \simeq 0.28 \pm 0.04$. This scaling is not inconsistent with the discussion of the origin of the alignment given in Section 3 in as much as a smoothing length defines a set of filaments and therefore picks out a halo mass scale corresponding to the haloes that are flowing/merging along these filaments. Note that in practice the dependence on smoothing is actually rather weak (expressed in terms of mass, we have $M_{\rm crit} \propto M^{0.27}_{\rm smooth}$ for filaments defined via the Skeleton and $M_{\rm crit} \propto M^{0.09}_{\rm smooth}$ if structure anisotropy is described via the shear tensor). Fig. D1 also shows that α depends on redshift. Note that as expected, the critical mass for the tidal tensor matches that of the filaments smoothed on a larger scale ($\simeq 7.5\,h^{-1}$ Mpc instead of 5). This confirms the idea that the potential is close to a smoother version of the density field.

D2 Non-linear mass evolution

It is of interest to compare the transition masses, $M^{\rm s}_{\rm crit}(z)$ and $M^{\rm t}_{\rm crit}(z)$, with the mass scale that tracks the development of non-linearity in structure formation. The variance of the density field smoothed on scale R obeys

$$\sigma^2(R,z) = D(z)^2 \int_0^\infty P(k) W^2(kR)\,{\rm d}^3k, \qquad \text{(D2)}$$

with $P(k)$ the power spectrum, and the top-hat filter defined by $W^2(x) = 9(\sin x/x - \cos x)^2/x^2$. The growth factor $D(z)$ is given by

$$D(z) = \frac{5}{2}\Omega_{\rm m} H_0^2 H(z) \int_z^\infty \frac{(1+z)\,{\rm d}z}{H(z)^3},$$

with $H(z) = H_0\sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda}$. Here H_0, $\Omega_{\rm m}$ and Ω_Λ are the Hubble constant, the DM and the dark energy density parameters at $z = 0$, respectively.

Fixing the level of (non-)linearity by the condition $\sigma(R(z), z) =$ constant implicitly defines the redshift evolution of the smoothing scale $R(z)$ (expressed in comoving Mpc) that maintains this level of (non-)linearity. This, in turn, corresponds to the mass scale

$$M_{\rm NL}(z) \equiv \frac{4}{3}\pi\bar{\rho}\,R(z)^3, \qquad \text{(D3)}$$

where $\bar{\rho}$ is the present-day average density of matter in the Universe.

For the matter-dominated CDM Universe with a scale-free power spectrum $P(k) \propto k^n$, $M_{\rm NL}(z) \propto (1+z)^{-6/(n+3)}$. In the Universe with realistic parameters, the redshift dependence of $M_{\rm NL}(z)$ is not a power law, both due to the influence of the Λ term that slows down the growth of the structure at low redshifts, and due to the steepening of the spectrum as one moves to smaller scale at high redshifts.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Animations. *Movie1.mov*: a very small subset of skeletons at different redshifts extracted from the Horizon 4π simulation, together with unit vectors showing the orientation of the spin of the corresponding dark matter halo with a mass above the transition mass. The spin is indeed perpendicular to the filament for these massive haloes.
Movie2.mov: time sequence evolution of gas tracer particles trapped in walls and filaments and coalescing to form a central galaxy.
Movie3.mov and Movie4.mov: same as *Movie2.mov* from another projection.
Movie5.mov: gas tracer particles around a forming galaxy; the ribbon-like filamentary structure feeding the central galaxy is clearly visible on this circular view.
Movie6.mov: the time line of a set of tracer particles is shown; each tracer particle is colour coded with a different colour. First the viewer rotates around the large-scale distribution of the time line tracer particle; later only the inner region and late times are shown.
Movie7.mov: tracking in towards a set of filaments, colour coded by tracer particle density at high redshift. The coplanar distribution of filaments is clearly seen, together with young galaxies which tend to sit perpendicular to the filament axes. The cold gas describes ribbons corresponding to the locus of the shocks which feed these young galaxies.
Movie8.mov: same tracking in at some intermediate redshift; some galaxies are in the process of merging.
Movie9.mov: same tracking in at some later redshift; the result of mergers corresponding to galaxies which have a spin perpendicular to the main axes of the filaments.
Movie10.mov: a zoom over the inner region of *Movie7.mov* which illustrates the walls of the large-scale structures, the shock-induced ribbons and a nice example of a galaxy formed with a spin parallel to its filament.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **452,** 3369–3393 (2015) doi:10.1093/mnras/stv1570

Spin alignments within the cosmic web: a theory of constrained tidal torques near filaments

Sandrine Codis,[1]⋆ Christophe Pichon[1,2] and Dmitry Pogosyan[3]

[1]*Institut d'Astrophysique de Paris, CNRS & UPMC, UMR 7095, 98 bis Boulevard Arago, F-75014 Paris, France*
[2]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3]*Department of Physics, University of Alberta, 11322-89 Avenue, Edmonton, Alberta T6G 2G7, Canada*

Accepted 2015 July 10. Received 2015 June 28; in original form 2015 April 23

ABSTRACT

The geometry of the cosmic web drives in part the spin acquisition of galaxies. This can be explained in a Lagrangian framework, by identifying the specific long-wavelength correlations within the primordial Gaussian random field (GRF), which are relevant to spin acquisition. Tidal torque theory is revisited in the context of such anisotropic environments, biased by the presence of a filament within a wall. The point process of filament-type saddles represents it most efficiently. The constrained misalignment between the tidal and the inertia tensors in the vicinity of filament-type saddles simply explains the distribution of spin directions. This misalignment implies in particular an *azimuthal* orientation for the spins of more massive galaxies and a spin *alignment* with the filament for less massive galaxies. This prediction is found to be in qualitative agreement with measurements in GRFs and *N*-body simulations. It relates the transition mass to the geometry of the saddle, and accordingly predicts its measured scaling with the mass of non-linearity. Implications for galaxy formation and weak lensing are briefly discussed, as is the dual theory of spin alignments in walls.

Key words: galaxies: evolution – galaxies: formation – galaxies: kinematics and dynamics – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

Modern simulations based on a well-established paradigm of cosmological structure formation predict a significant connection between the geometry and dynamics of the large-scale structure on the one hand, and the evolution of the physical properties of forming galaxies on the other. Key questions formulated decades ago are nevertheless not fully answered. What are the main processes which determine the morphology of galaxies? What is the role played by angular momentum in shaping them?

Pichon et al. (2011) have suggested that the large-scale coherence of the inflow, inherited from the low-density cosmic web, explains why cold flows are so efficient at producing thin high-redshift discs from the inside out (see also Stewart et al. 2013; Prieto et al. 2015; Laigle et al. 2015). On the scale of a given gravitational patch, gas is expelled from adjacent voids, towards sheets and filaments forming at their boundaries. Within these sheets/filaments, the gas shocks and radiatively loses its energy before streaming towards the nodal points of the cosmic network. In the process, it advects angular momentum, hereby seemingly driving the morphology of galaxies (bulge or disc). The evolution of the Hubble sequence in such a scenario is therefore at least in part initially driven by the geometry of the cosmic web. As a consequence, the distribution of the properties of galaxies measured relative to their cosmic web environment should reflect such a process. In particular, the spin distribution of galaxies should display a preferred mass-dependent orientation relative to the cosmic web.

Both numerical (e.g. Aragón-Calvo et al. 2007; Hahn et al. 2007; Sousbie et al. 2008; Paz, Stasyszyn & Padilla 2008; Zhang et al. 2009; Codis et al. 2012; Libeskind et al. 2013; Aragón-Calvo & Yang 2014; Dubois et al. 2014), and observational evidence (e.g. Tempel et al. 2013) have recently supported this scenario. In parallel, much analytical (e.g. Catelan, Kamionkowski & Blandford 2001; Hirata & Seljak 2004), numerical (e.g. Heavens, Refregier & Heymans 2000; Croft & Metzler 2000; Schneider & Bridle 2010; Schneider, Frenk & Cole 2012; Joachimi et al. 2013b; Codis et al. 2015; Tenneti et al. 2015) and observational (e.g. Brown et al. 2002; Lee & Pen 2002; Bernstein & Norberg 2002; Heymans et al. 2004; Hirata et al. 2004, 2007; Hirata & Seljak 2004; Mandelbaum et al. 2006, 2011; Joachimi et al. 2011, 2013a) efforts have been invested to control the level of intrinsic alignments of galaxies as a potential source of systematic errors in weak gravitational lensing measurements. Such alignments are believed to be a worrisome source of systematics of the future generation of lensing surveys like Euclid or Large Synoptic Survey Telescope. It is therefore of interest to understand from first principles why such intrinsic alignments arise, so as to possibly temper their effects.

⋆ E-mail: codis@iap.fr

Hence we should try and refine a theoretical framework to study the dynamical influence of filaments on galactic scales, via an extension of the peak theory to the truly three-dimensional anisotropic geometry of the circum-galactic medium, and amend the standard galaxy formation model to account for this anisotropy. Towards this end, we will develop here a filament version of an anisotropic 'peak-background-split' formalism, i.e. make use of the fact that walls and filaments are the interference patterns of primordial fluctuations on large scales, and induce a corresponding anisotropic boost in overdensity. Indeed, filaments feeding galaxies with cold gas are themselves embedded in larger scale walls imprinting their global geometry (Danovich et al. 2012; Dubois et al. 2012).

On top of these modes, constructive interferences of high-frequency modes produce peaks which thus get a boost in density that allows them to pass the critical threshold necessary to decouple from the overall expansion of the Universe, as envisioned in the spherical collapse model (Gunn & Gott 1972). This well-known biased clustering effect has been invoked to justify the clustering of galaxies around the nodes of the cosmic web (White, Tully & Davis 1988). It also explains why galaxies form in filaments: in walls alone, the actual density boost is typically not sufficiently large to trigger galaxy formation. The main nodes of the cosmic web are where galaxies migrate, not where they form. They thus inherit the anisotropy of their birth place as spin orientation. During migration, they may collide with other galaxies/haloes and erase part of their birth heritage when converting orbital momentum into spin via merger (e.g. Codis et al. 2012). Tidal torque theory should therefore be revisited to account for the anisotropy of this filamentary environment on various scales in order to model primordial *and* secondary spin acquisition.

In this paper, we will quantify and model the intrinsically 3D geometry of galactic spins while accounting for the geometry of saddle points of the density field. Indeed, saddle points define an anisotropic *point process* which accounts for the presence of filaments embedded in walls (Pogosyan et al. 1998), two critical ingredient in shaping the spins of galaxies.

Taking them into account will in particular allow us to predict the biased geometry of the tidal field in the vicinity of saddle points. This can be formalized using the two-point joint probability of the gravitational potential field and its first to fourth derivatives and imposing a saddle-point constraint. For Gaussian (or quasi-Gaussian) fields these two-point functions are within reach from first principle (Bardeen et al. 1986). A proper account of the anisotropy of the environment in this context will allow us to demonstrate why the spin of the forming galaxies field are first aligned with the filament's direction.

We will also show that massive galaxies will have their spin preferentially along the azimuthal direction. While relying on a straightforward extension of Press–Schechter's theory, we will predict the corresponding transition mass' scaling with the (redshift-dependent) mass of non-linearity, while relying on the so-called cloud-in-cloud problem applied to the filament-background split.

The paper is organized as follows. Section 2 qualitatively presents the basis of the physical process at work in aligning the spin of dark haloes relative to the cosmic web. Section 3 then presents the expected Lagrangian spin distribution near filaments, assuming cylindrical symmetry, and explains the observed mass transition while carrying a multiscale analysis of the fate of collapsing haloes in the vicinity of 2D saddle points. Section 4 revisits this distribution in three dimensions for realistic typical 3D saddle points. Section 5 investigates the predictions of the theory using Gaussian random field (GRF) and *N*-body simulations, while we finally conclude in Section 6. Appendix A discusses possible limitations and extensions of this work. Appendix B presents the dual theory for spin alignment near wall saddles. Finally, Appendix C gathers some technical complements.

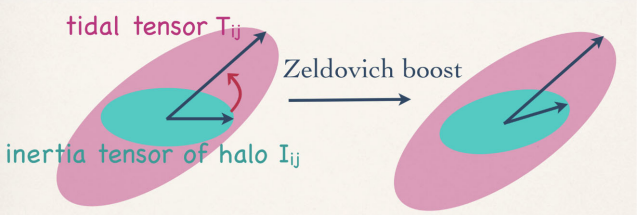


Figure 1. Spin acquisition by tidal torquing. At linear order, the misalignment between the inertia tensor of the proto-object and the surrounding tidal tensor induces an inhomogeneous Zel'dovich boost which corresponds to the acquisition of a net intrinsic angular momentum in Eulerian space.

2 TIDAL TORQUING NEAR A SADDLE

Before presenting analytical estimates for the expected spins near filament in two and three dimensions and their transition mass, let us discuss qualitatively what underpins the corresponding theory.

2.1 Spin acquisition by tidal torquing

In the standard paradigm of galaxy formation, protogalaxies acquire their spin[1] by tidal torquing coming from the surrounding matter distribution (Hoyle 1949; Peebles 1969; Doroshkevich 1970; White 1984; Catelan & Theuns 1996; Crittenden et al. 2001). At linear order, this spin is acquired gradually until the time of maximal extension (before collapse) and is proportional to the misalignment between the inertia tensor of the protogalaxy and the surrounding tidal tensor (see Schaefer 2009, for a review)

$$L_i = \sum_{j,k,l} a^2(t)\dot{D}_+(t)\epsilon_{ijk} I_{jl} T_{lk}\,, \tag{1}$$

where $a(t)$ is the scale factor, D_+ the growth factor, T_{ij} the tidal tensor (detraced Hessian of the gravitational potential), I_{ij} the protogalactic inertia tensor (only its traceless part, $\bar{I}_{ij}$ contributes to the spin). As this work focuses on the spin direction, the factor $a^2(t)D_+(t)$ will henceforth be dropped for brevity. This process of spin acquisition by tidal torquing is illustrated on Fig. 1.

In the Lagrangian picture, I_{ij} is the moment of inertia of a uniform mass distribution within the Lagrangian image of the halo, while T_{ij} is the tidal tensor averaged within the same image. Thus, to rigorously determine the spin of a halo, one must know the area from which matter is assembled, beyond the spherical approximation. While this can be determined in numerical experiments, theoretically we do not have the knowledge of the exact boundary of a protohalo. As such, one inevitably has to introduce an approximate proxy for the moment of inertia (and an approximation for how the tidal field is averaged over that region).

The most natural approach is to consider that protohaloes form around an elliptical peak in the initial density and approximate its Lagrangian boundary with the elliptical surface where the overdensity drops to zero. This leads to the following approximation for the

[1] Note that in this paper we will call interchangeably 'spin' or 'angular momentum' the intrinsic angular momentum of (proto)haloes.

traceless part of the inertia tensor (e.g. Schäfer & Merkel 2012, see also equations A2–A4)

$$\overline{I}_{ij} = \frac{2}{5}\nu\sigma_2 M \overline{H}^{-1}_{ij} = \frac{2}{5}\nu\sigma_2 \frac{M}{\det \boldsymbol{H}} \widetilde{H}_{ij} \,, \tag{2}$$

where $\overline{H}^{-1}_{ij}$ is the traceless part of the inverse Hessian of the density field, $H_{ij} = \partial_i \partial_j \delta$, ν is the overdensity at the peak and M is the mass of the protohalo. In the second form we explicitly presented the inverse Hessian via the (detraced) matrix of the Hessian minors, $\widetilde{H}_{ij}$. While $\widetilde{H}_{ij}$ is a simple polynomial in second derivatives of the density, $M/\det \boldsymbol{H}$ is not, which is the source of most technical difficulties when statistical studies of the spin are attempted.

Let us point at the following considerations to bypass these difficulties. First, the supplementary condition for the halo to be at a peak of the density yields an extra $\det \boldsymbol{H}$ factor in all statistical measures (see e.g. Bardeen et al. 1986). This factor exactly cancels the determinant in the denominator. Secondly, all quantities in equation (2) are computed after the density field is smoothed at a particular scale R_h which sets the corresponding mass scale. Therefore, it is more appropriate to apply equation (2) to haloes at *fixed* mass M, determined by that smoothing. Hence, we could argue for the proxy $\overline{I}_{ij} \propto \nu \widetilde{H}_{ij}$ for the moment of inertia for haloes of a given fixed mass, where the change in mass is reflected in the corresponding change in the smoothing scale. In two dimensions, we show in Appendix A that this multiscale approximation gives qualitatively the same statistical results as just using H_{ij} as a proxy. While this approximation is relatively simple, since we are only concerned with the direction of the spin, we will now go one step further and use throughout this paper the Hessian as a proxy for the inertia tensor, even in three dimensions. Indeed, I_{ij}, $\widetilde{H}_{ij}$ and H_{ij} share the same eigen-directions (Catelan & Theuns 1996; Schäfer & Merkel 2012), so we *define* the spin for the rest of the paper as

$$s_i \equiv \sum_{j,k,l} \epsilon_{ijk} H_{jl} T_{lk} \,. \tag{3}$$

The vector field s_i is then quadratic in the successive derivatives of the potential: its (possibly constrained) expectation can therefore be computed for GRFs. This approximation is further discussed in Appendix A. Note that equation (3) improves upon simple parametrizations of the mean misalignment between inertia and tidal tensors (see e.g. Lee & Pen 2000; Crittenden et al. 2001) by *ab initio* explicitly taking into account the correlations between both tensors.

2.2 Geometry of the cosmic web

Galaxies are not forming everywhere but preferentially in filaments and nodes which define the so-called cosmic web (Klypin & Shandarin 1993; Bond, Kofman & Pogosyan 1996). The origin of these structures lies in the asymmetries of the initial GRF describing the primordial universe, amplified by gravitational collapse (Zel'dovich 1970). The presence of such large-scale structure (walls, filaments, nodes) induces local preferred directions for both the tidal tensor and the inertia tensor of forming objects which will eventually turn into preferred alignments of the spin w.r.t the cosmic web. It is therefore of interest to understand what is the expected spin direction predicted by equation (3) given the presence of a typical filament nearby. As a filament is typically the field line that joins two maxima of the density field through a filament-type saddle point (where the gradient is null and the density Hessian has two negative eigenvalues), we choose to study in this paper the expected spin direction of proto-objects *in the vicinity of a filament-type saddle point* with a given geometry (which imposes the direction of the filament and the wall). Fig. 2 illustrates the geometry of filaments near peaks and saddles in a 2D Gaussian field.

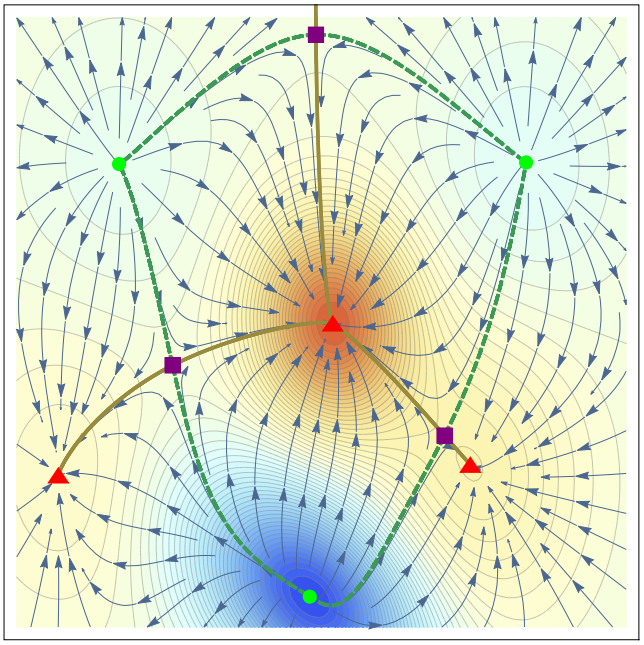


Figure 2. On top of the density contours (from dark blue to dark red), the three (red triangle) maxima (resp. the three (green points) minima) are connected by the crest lines (in solid gold, resp. the through lines in dashed green) which intersect through saddles points (purple squares). The blue arrows represent stream lines of the gradient flow. Throughout this paper, we will assume that the geometry of the regions of intermediate densities are set by the shape of the (purple) saddles.

2.3 Constrained tidal torque theory in a nutshell

2.3.1 *Spin alignments and flips*

It has been shown in simulations (among others Bailin & Steinmetz 2005; Aragón-Calvo et al. 2007; Paz et al. 2008; Zhang et al. 2009; Codis et al. 2012; Libeskind et al. 2013; Forero-Romero, Contreras & Padilla 2014) that the spin of dark haloes is correlated to the direction of the filaments of the cosmic web in a mass-dependent way. The alignment between the spin and the closest filament increases with mass until a mass of maximum alignment (Laigle et al. 2015) that we call here critical mass. As mass increases, the direction of the spin becomes less aligned with the filament before becoming perpendicular to it (Codis et al. 2012). This transition – from aligned to perpendicular – occurs at a mass that we call here the transition mass.

This paper will claim that the critical mass is directly related to the size of the quadrant of coherent angular momentum imposed by the tides of the saddle point (which are effectively the Lagrangian counter parts of the quadrant of vorticity found in Laigle et al. 2015). This mass can be captured using a cylindrical model that would correspond to the plane perpendicular to the filament at the saddle point (which amounts to assuming an infinitely long filament). This 2D toy model (see Section 3 below) shows that near a 2D peak (i.e. near an infinitely long 3D filament), the quadrupolar structure seen in simulation naturally arises in a Lagrangian framework. We investigate the size of that quadrants and shows that it qualitatively predicts the right critical mass.

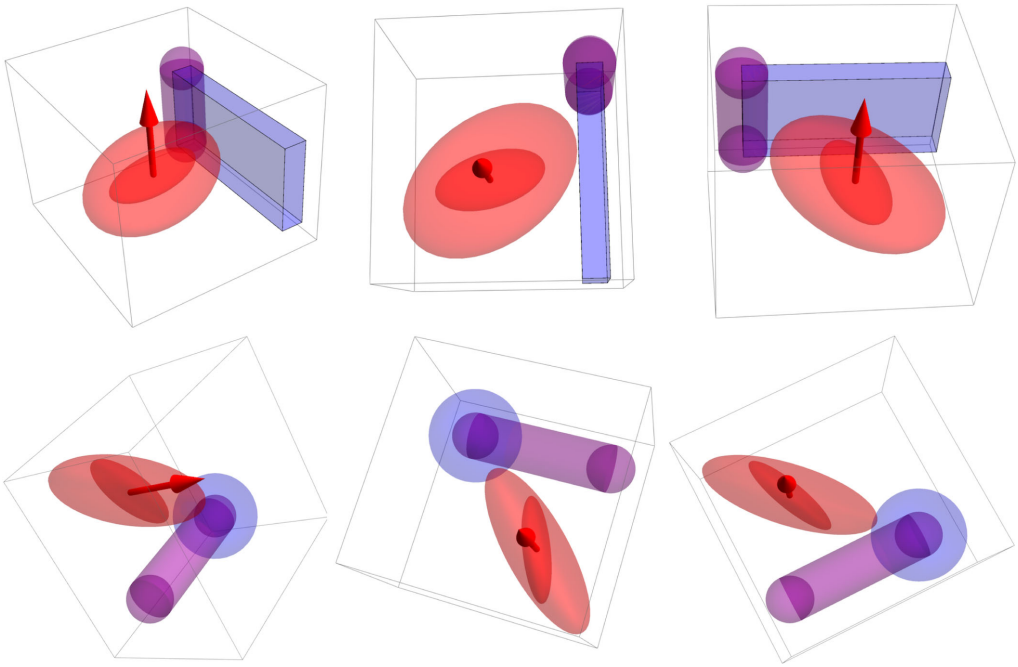


Figure 3. Sketch of main differential alignment between halo shapes and tidal tensor responsible for $\boldsymbol{e}_z$- and $\boldsymbol{e}_\phi$-component of momentum. Top: the two tensors in light and dark red, end up being misaligned as they feel differently the neighbouring wall (blue) and filament (purple), inducing a spin parallel to the filament (red arrow). Three projections are shown for clarity. Bottom: correspondingly, the differential pull from the filament (purple) and the density gradient towards the peak (blue) generates a spin (red arrow) along the azimuthal direction. By symmetry, the other peak(s) on the other side of the saddle point will spin-up massive haloes in the opposite direction.

The second stage of accretion, that flips the spin of more massive haloes from aligned to perpendicular to the filaments, requires a 3D analysis (see Section 4). It is shown that indeed small haloes that form close to the saddle point, acquire spin along the filaments while more massive haloes that form further from the saddle (i.e. closer to the peaks/nodes) acquire a spin perpendicular to the filaments (while accreting smaller haloes). The transition mass will be predicted as a function of redshift and shown to agree with measurements in simulations.

2.3.2 *The premises of anisotropic tidal torque theory*

Let us present here an outline of the extension of tidal torque theory (TTT) within the context of a peak (or saddle) background split. Given the anisotropically triaxial saddle constraint, we will argue that the misalignment between the tidal tensor and the Hessian of the density field simply explains the transverse and longitudinal anti-symmetric geometry of angular momentum distribution in their vicinity. It arises because the two tensors probe different scales: given their relative correlation lengths, the Hessian probes more directly its closest neighbourhood, while the tidal field, somewhat larger scales.

Within the plane of the saddle point perpendicular to the filament axis (the mid-plane hereafter), the dominant wall (corresponding to the longer axis of the cross-section of the saddle point) will re-orient more the Hessian than the tidal tensor, which also feels the denser, but typically further away saddle point, see Fig. 3, top panels. This net misalignment will induce a spin perpendicular to that plane, i.e. along the filament. This effect will produce a quadrupolar, anti-symmetric distribution of the longitudinal component of the angular momentum which will be strongest at some four points, not far off-axis. Beyond a couple of correlation lengths away from those four points, the effect of the tidal field induced by the saddle point will subside, as both tensors become more spherical.

Conversely, in planes containing the filament, e.g. containing the main wall, a similar process will misalign both tensors. This time, the two anisotropic features differentially pulling the tensors are the filament on the one hand, and the density gradient towards the peak on the other. The net effect of the corresponding misalignment will be to also spin-up haloes perpendicular to that plane, along the azimuthal direction, see Fig. 3, bottom panels. By symmetry, the anti-clockwise tidal spin will be generated on the other side of the saddle point.

Hence, the geometry of angular momentum near filament-saddle points is the following: it is aligned with the filament in the median plane (within four anti-symmetric quadrants), and (anti-)aligned with the azimuthal direction away from that plane. The stronger the triaxiality the stronger the amplitude. Conversely, if the saddle point becomes degenerate in one or two directions, the component of the angular momentum in the corresponding direction will vanish. For instance, a saddle point in the middle of a very long filament will only display alignment with that filament axis, with no azimuthal component. For a typical triaxial configuration, two pairs of four points define the loci of maximal longitudinal and azimuthal spin.

2.3.3 *Geometry of spin flip*

Fig. 4 gives a more quantitative account of the geometry of the tidal field around a given saddle point embedded in a given dominant wall. We consider here the angular momentum distribution near a filament-saddle point, *S*. It is assumed that the three eigenvalues of the density are such that the filament going through this saddle point is along the vertical axis and that the other two eigenvalues are different, reflecting the presence of a dominant wall, in the *Sxz* plane,

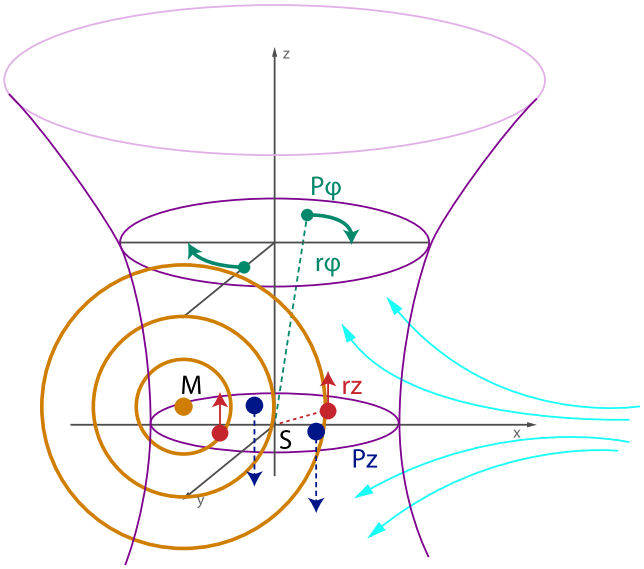


Figure 4. Qualitative geometry of the angular momentum distribution near a elliptical saddle point S (see also Fig. 12). The shape of a given triaxial isodensity is shown in purple, together with two cross-sections, resp. in the Sxy mid-plane and in a plane containing the maxima of transverse angular momentum. The velocity flow in the Sxz plane is shown in cyan. The locus of the Lagrangian extent of haloes is shown in orange concentric spheres centred on M. In the Sxy plane, the four points, P_z correspond to the maxima of the modulus of angular momentum. They point, respectively, along $\boldsymbol{e}_z$ in the first and third quadrants (in red), and along $-\boldsymbol{e}_z$ in the second and fourth quadrant (in blue). Conversely, the four points, P_ϕ, correspond to the maxima of the alignment of $\boldsymbol{s}$ along $\boldsymbol{e}_\phi$. Only two (above the mid-plane) are shown. As the orange sphere capture more than one quadrant, the z-component of $\boldsymbol{L}$ subside, as it encompasses more of the neighbourhood of P_ϕ, and its $\boldsymbol{e}_\phi$-component increases. In this Lagrangian framework, the spin flip as a function of mass is a direct consequence of the geometry of the tidal field imposed by the saddle point.

in which the filament is embedded. The shape of a given triaxial isodensity is shown in purple, together with two cross-sections, resp. in the Sxy mid-plane and in a plane containing the maxima of transverse angular momentum. As we will demonstrate later, the spin is mostly confined in the neighbourhood of the Sz-axis, up to a couple of correlation length of the density. It would in fact vanish, should the saddle become isotropic. In the Sxy plane, we identify four quadrants corresponding to regions in which the spin is parallel to the filament. Within theses quadrant, the spin point, respectively, along $\boldsymbol{e}_z$ in the first and third quadrants, and along $-\boldsymbol{e}_z$ in the second and fourth quadrants. By symmetry, the spin has to vanish along Sx and Sy.

2.3.4 *Towards a transition mass?*

The twisted geometry of the spin near the saddle point also allows us to identify the Lagrangian transition mass corresponding to the alignment of dark matter (DM) haloes' spin relative to the direction of their neighbouring filament. Let us first consider the Lagrangian counterpart of a low-mass halo and assume it lies near the median plane. it will typically fall into one of the quadrant corresponding to an orientation of the spin parallel to the filament axis. Now consider a halo of larger Lagrangian extent. As long as its size is smaller than the typical size of a quadrant (which will be defined more precisely below) the alignment increases, until it overextends the quadrant. As it does, two things happen (i) it will start capturing tides from the next quadrant, which would anti-align it; as the Lagrangian patch radius increases more, it reaches a size comparable to the whole tidal region of influence of the saddle point. It then encompasses both the clockwise and anti-clockwise azimuthal regions, and add up to a net momentum of null amplitude. (ii) it will start capturing the effect of the azimuthal tide, hence inducing a spin flip. Depending on the ratio of the eigenvalues of the Hessian, the two might be concurrent or not. In parallel, as the radius increases, the patch collects the mean potential gradient which defines the Zel'dovich boost which will drive it away from the neighbourhood of the saddle point. The above description clearly accounts for the influence of only one saddle point. As we consider regions further away from that saddle, we should account for the influence of other critical points, as discussed in Section 5.

We have up to now considered a patch centred near the mid-plane close to the saddle point. Indeed, typically, in the peak-background-split framework, such patches will collapse preferentially where the density is boosted, that is within the wall containing the filament, close to the filament. The rarer (more massive) haloes will form in turn in the denser regions, away from the saddle point, along the filament, while the more common lighter haloes will form everywhere and in particular near the saddle point. The former will have a spin perpendicular to the filament. The latter will have a spin parallel to the filament. The relative number of light to small haloes will depend on curvilinear coordinate along the filament because consumption is important: object above the transition mass have swallowed their lighter parents. At a given redshift, the left overs will decide what matters. This effect is the anisotropic version of the well-known cloud-in-cloud problem.

2.3.5 *Lagrangian dynamics of spin flip*

In order to understand how a *given* halo flips, let us split the original Lagrangian patch in two concentric shells. The inner shell will correspond to the Lagrangian extent of the halo as it initially forms, while the outer shell will correspond to secondary infall. The reasoning presented in Section 2.3.2 can be applied independently to both the inner and outer shells, and we would typically conclude that the outer shell would be more likely to have its spin perpendicular to the filament axis. It follows that, as far as this halo is concerned, it will undergo a spin flip as it moves towards the core of the filament and away from the saddle. This process will also correspond to an acquired net *helicity* for the secondary infall, which will last as long as the transverse anisotropy of the saddle point correlates the local tidal field. In effect, this consistent helicity will build up the spin of the forming galaxy via secondary infall as it drifts, up to the point where mergers will re-orient the direction of its spin. Hence, this constructive build-up of disc should only last so long as the galaxy drifts within the high-helicity region. Note that the transverse motion will correspond to the halo entering the vortex rich caustic corresponding to the multiflow region near the filament, so that this Lagrangian description remains fully consistent with the Eulerian discussion given in Laigle et al. (2015). We can anticipate that the longitudinal motion generates azimuthal vortices as well.

The scenario described in this section can be formalized at two levels. First, within the framework of constrained random fields, one can compute the expected geometry of the spin configuration near a given saddle. This will yield a map of the mean alignment between spin and filament in the vicinity of the saddle point. We will then marginalize over the expected distribution of such saddles, and model correspondingly the evolution of the expected mass of

dark haloes around the filament. This will allow us to recover the numerically measured mass transition for spin flip. We may also test the mass-dependent alignment w.r.t. $\boldsymbol{e}_\phi$ in GRFs and N-body simulations. For the sake of clarity, we will proceed in two steps: first, while assuming cylindrical symmetry we will compute the expected spin distribution within the most likely cross-section of a filament of infinite extend (Section 3); then we will compute this expectation around the most likely 3D saddle point (Section 4).

3 SPIN ALONG INFINITE FILAMENT

Let us first start while assuming that the filament is of infinite extent, so that we can restrict ourselves to cylindrical symmetry in two dimensions. This is of interest as the angular momentum is then along the filament axis by symmetry and its derivation in the context of TTT is much simpler. It captures already in part the mass transition, in as much as we can define the mean extension of a given quadrant of momentum with a given polarity. In this context, it is of interest to study the spin geometry in the median plane, i.e. in the vicinity of a 2D peak. This 2D spin is along the filament, and will be denoted s_z in what follows.

3.1 Shape of the spin distribution near filaments

Under the assumption that the *direction* of the spin along the z direction is well represented by the fully anti-symmetric (Levi Civita) contraction of the tidal tensor and density Hessian given by equation (3) (e.g. Schäfer & Merkel 2012), it becomes a quadratic function of the second and fourth derivatives of the potential. As such, it becomes possible to compute expectations of it subject to its relative position to a peak with a given geometry (which would correspond to the cross-section of the filament in the mid-plane). Note that, as mentioned in Section 2.1, standard TTT relies, more correctly, on the inertia tensor in place of the Hessian. Even though they have inverse curvature of each other, their set of eigen-directions are locally the same, so we expect the induced spin *direction* – which is the focus of this paper, to be the same, so long as the inertia tensor is well described by its local Taylor expansion.

3.1.1 *Constrained joint PDF near peak*

Any matrix of second derivatives f_{ij} – rescaled so that $\langle(\Delta f)^2\rangle = 1$ – can be decomposed into its trace Δf, and its detraced components in the frame of the separation

$$f^+ = (f_{11} - f_{22})/2\,, \quad f^\times = f_{12}. \tag{4}$$

Then all the correlations between two such matrices, f_{ij} and g_{ij} can be decomposed irreducibly as follows. Let us call $\xi_{fg}^{\Delta\Delta}$, $\xi_{fg}^{\Delta+}$ and $\xi_{fg}^{\times\times}$ the correlation functions in the frame of the separation (which is the first coordinate here) between the second derivatives of the field f and g separated by a distance r:

$$\begin{aligned}
\xi_{fg}^{\Delta\Delta}(r) &= \langle \Delta f\, \Delta g\rangle\,,\\
\xi_{fg}^{\Delta+}(r) &= \left\langle \Delta f g^+\right\rangle,\\
\xi_{fg}^{\times\times}(r) &= \left\langle f^\times g^\times\right\rangle.
\end{aligned} \tag{5}$$

All other correlations are trivially expressed in terms of the above as

$$\left\langle f^\times \Delta g\right\rangle = 0,\ \left\langle f^+ g^\times\right\rangle = 0,$$

$$\left\langle f^+ g^+\right\rangle = \frac{1}{4}\xi_{fg}^{\Delta\Delta}(r) - \xi_{fg}^{\times\times}(r)\,. \tag{6}$$

Here, we consider two such fields, namely the gravitational potential Φ and the density δ. In the following, these two fields and their first and second derivatives are assumed to be rescaled by their variance $\sigma_0^2 = \langle\Phi^2\rangle$, $\sigma_1^2 = \langle(\nabla\Phi)^2\rangle$, $\sigma_2^2 = \langle(\delta = \Delta\Phi)^2\rangle$, $\sigma_3^2 = \langle(\nabla\delta)^2\rangle$ and $\sigma_4^2 = \langle(\Delta\delta)^2\rangle$. The shape parameter of the density field is defined as

$$\gamma = \sigma_3^2/(\sigma_2\sigma_4). \tag{7}$$

The rescaled potential and density will be denoted by ϕ and x and the rescaled first and second derivatives by ϕ_i, x_i and ϕ_{ij}, x_{ij}.

Let us gather the first and second derivatives of the gravitational field and the first and second derivatives of the density in a vector denoted by $\boldsymbol{X}$ spatially located in $\boldsymbol{r}_X$ and $\boldsymbol{Y}$ located in $\boldsymbol{r}_Y$. The Gaussian joint probability distribution function (PDF) of $\boldsymbol{X}$ and $\boldsymbol{Y}$ at the two given locations ($\boldsymbol{r}_X$ and $\boldsymbol{r}_Y$ separated by a distance $r = |\boldsymbol{r}_X - \boldsymbol{r}_Y|$) obeys

$$\mathcal{P}(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{\sqrt{\det|2\pi\,\mathbf{C}|}}\exp\left(-\frac{1}{2}\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix}^{\mathrm{T}}\cdot\mathbf{C}^{-1}\cdot\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix}\right), \tag{8}$$

where $\mathbf{C}_0 \equiv \langle\boldsymbol{X}\cdot\boldsymbol{X}^{\mathrm{T}}\rangle$, $\mathbf{C}_\gamma \equiv \langle\boldsymbol{X}\cdot\boldsymbol{Y}^{\mathrm{T}}\rangle$ and

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_0 & \mathbf{C}_\gamma\\ \mathbf{C}_\gamma^{\mathrm{T}} & \mathbf{C}_0\end{bmatrix}.$$

All these quantities depend on the separation vector $\boldsymbol{r}$ only because of statistical homogeneity. This PDF is sufficient to compute the expectation of any quantity involving derivatives of the potential and the density up to second order. All the coefficients can easily be computed from the power spectrum of the potential

$$\left\langle \partial_1^{i_1}\partial_2^{i_2}\phi,\, \partial_1^{j_1}\partial_2^{j_2}\phi\right\rangle = \int_0^\infty\int_0^{2\pi}\mathrm{d}\theta\ \mathrm{d}k\ P_k(k)\exp(\imath k\, r\cos\theta)$$
$$\imath^{i_1+i_2}(-\imath)^{j_1+j_2}(\cos\theta)^{i_1+j_1}(\sin\theta)^{i_2+j_2}\frac{k^{i_1+i_2+j_1+j_2+1}}{\sigma_{i_1+i_2}\sigma_{j_1+j_2}}, \tag{9}$$

and

$$\sigma_n^2 = \int_0^\infty\int_0^{2\pi}\mathrm{d}\theta\ \mathrm{d}k\ P_k(k)k^{2n+1}\,,$$

where the power spectrum of the potential $P_k(k)$ can include a filter function on a given scale. In this work, we use a Gaussian filter defined in Fourier space by

$$W_G(\boldsymbol{k}, R) = \frac{1}{(2\pi)^{3/2}}\exp\left(\frac{-k^2R^2}{2}\right). \tag{10}$$

For instance, the one-point covariance matrix for ($\Delta\phi$, ϕ^+, $\phi^\times$, Δx, x^+, $x^\times$) at a given point simply reads

$$\mathbf{C}_{02} = \begin{pmatrix}
1 & 0 & 0 & -\gamma & 0 & 0\\
0 & 1/8 & 0 & 0 & -\gamma/8 & 0\\
0 & 0 & 1/8 & 0 & 0 & -\gamma/8\\
-\gamma & 0 & 0 & 1 & 0 & 0\\
0 & -\gamma/8 & 0 & 0 & 1/8 & 0\\
0 & 0 & -\gamma/8 & 0 & 0 & 1/8
\end{pmatrix},$$

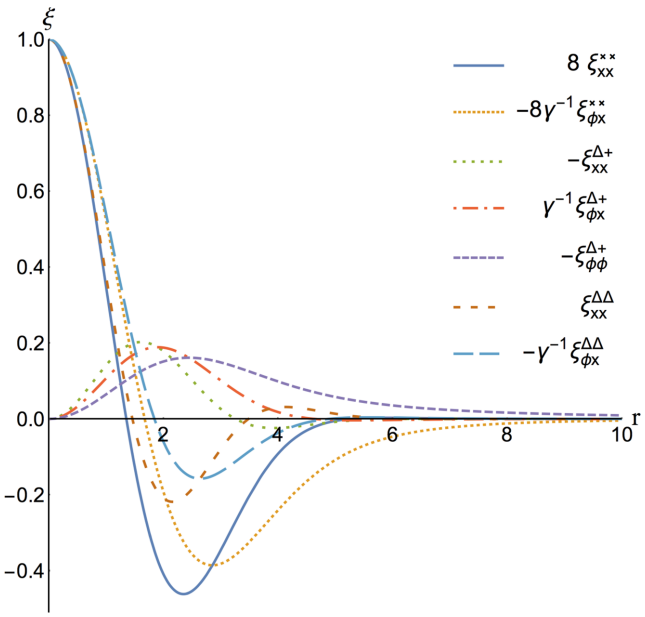


Figure 5. Two-point correlation functions as a function of the separation r in units of the smoothing length for a power-law 2D power spectrum with spectral index $n = -1/2$, i.e. $\gamma = \sqrt{3/7}$. For aesthetic purpose, these functions have been rescaled by their value in $r = 0$ as explicitly written in the legend.

where $\gamma = \sqrt{(n+2)/(n+4)}$ for a scale-invariant density power spectrum with spectral index n (i.e. $n - 4$ for the potential). Note that the first derivatives of the density and the potential fields are decorrelated from the second derivatives meaning that

$$\mathbf{C}_0 = \begin{pmatrix} \mathbf{C}_{01} & 0 \\ 0 & \mathbf{C}_{02} \end{pmatrix},$$

where $\mathbf{C}_{01}$ is the one-point covariance matrix of the gradients of the potential and the density fields $(\phi_1, \phi_2, x_1, x_2)$ as a function of $\gamma' = \sigma_2^2/\sigma_1/\sigma_3$

$$\mathbf{C}_{01} = \begin{pmatrix} 1/2 & 0 & -\gamma'/2 & 0 \\ 0 & 1/2 & 0 & -\gamma'/2 \\ -\gamma'/2 & 0 & 1/2 & 0 \\ 0 & -\gamma'/2 & 0 & 1/2 \end{pmatrix}.$$

The two-point covariance matrix, $\mathbf{C}_\gamma$ can be similarly derived. In particular, its restriction to the second derivatives of the density and the potential fields can be written as a function of the nine ξ functions defined in equation (5) (for $fg = \phi\phi, \phi x, xx$) (see Fig. 5 and Appendix C2):

$$\begin{pmatrix} \xi^{\Delta\Delta}_{\phi\phi} & \xi^{\Delta+}_{\phi\phi} & 0 & \xi^{\Delta\Delta}_{\phi x} & \xi^{\Delta+}_{\phi x} & 0 \\ \xi^{\Delta+}_{\phi\phi} & \frac{\xi^{\Delta\Delta}_{\phi\phi}}{4} - \xi^{\times\times}_{\phi\phi} & 0 & \xi^{\Delta+}_{\phi x} & \frac{\xi^{\Delta\Delta}_{\phi x}}{4} - \xi^{\times\times}_{\phi x} & 0 \\ 0 & 0 & \xi^{\times\times}_{\phi\phi} & 0 & 0 & \xi^{\times\times}_{\phi x} \\ \xi^{\Delta\Delta}_{\phi x} & \xi^{\Delta+}_{\phi x} & 0 & \xi^{\Delta\Delta}_{xx} & \xi^{\Delta+}_{xx} & 0 \\ \xi^{\Delta+}_{\phi x} & \frac{\xi^{\Delta\Delta}_{\phi x}}{4} - \xi^{\times\times}_{\phi x} & 0 & \xi^{\Delta+}_{xx} & \frac{\xi^{\Delta\Delta}_{xx}}{4} - \xi^{\times\times}_{xx} & 0 \\ 0 & 0 & \xi^{\times\times}_{\phi x} & 0 & 0 & \xi^{\times\times}_{xx} \end{pmatrix}$$

Once the joint PDF given by equation (8) is known, it is straightforward to compute conditional PDFs (in particular subject to a critical point constraint $\mathcal{C}(\mathrm{crit}) = |\det(x_{ij})|\delta_{\mathrm{D}}(x_i)$). Given the conditionals, simple algebra then yield the conditional density and spin. More specifically, relying on Bayes theorem, the conditional can be expressed in terms of the joint PDF – equation (8) – as

$$\mathcal{P}(\boldsymbol{X}|\boldsymbol{Y}, \mathrm{pk}) = \frac{\mathcal{P}(\boldsymbol{X}, \boldsymbol{Y}, \mathrm{pk})}{\mathcal{P}(\boldsymbol{Y}, \mathrm{pk})},$$

where

$$\mathcal{P}(\boldsymbol{Y}, \mathrm{pk}) = \int \mathrm{d}\boldsymbol{Y} \mathcal{P}(\boldsymbol{X}, \boldsymbol{Y}) \mathcal{C}(\mathrm{pk})$$

is the marginal distribution describing the likelihood of a given peak, pk (the transverse cross-section of an infinite filament) with a given geometry. Once the conditional, $\mathcal{P}(\boldsymbol{X}|\boldsymbol{Y}, \mathrm{pk})$ is known, it is straightforward[2] to compute the expectation of any function, $f(\boldsymbol{X})$ as

$$\langle f(\boldsymbol{X})|\mathrm{pk}\rangle = \int \mathrm{d}\boldsymbol{X}\, \mathcal{P}(\boldsymbol{X}|\mathrm{pk}) f(\boldsymbol{X}), \tag{11}$$

which, when $f(\boldsymbol{X})$ is multinomial in the components of $\boldsymbol{X}$ can be carried out analytically. In the following, we will consider in turn functions which are indeed algebraic function of $\boldsymbol{X}$.

3.1.2 *Constrained density maps*

From equation (8), given a contrast ν and a geometry for the saddle (or any critical point) defined by $\kappa = \lambda_1 - \lambda_2$, $I_1 = \lambda_1 + \lambda_2$ (where $\lambda_1 > \lambda_2$ are the two eigenvalues of the Hessian of the density field $\boldsymbol{H}$ – both negative for a peak), the mean density contrast, $\delta_{\mathrm{ext}} = \langle\delta|\mathrm{ext}\rangle$, (in units of σ_2) around the corresponding critical point can be analytically computed

$$\delta(\boldsymbol{r}|\mathrm{ext}, \kappa, I_1, \nu) = \frac{I_1\left(\xi^{\Delta\Delta}_{\phi x} + \gamma\xi^{\Delta\Delta}_{\phi\phi}\right) + \nu\left(\xi^{\Delta\Delta}_{\phi\phi} + \gamma\xi^{\Delta\Delta}_{\phi x}\right)}{1-\gamma^2} + 4(\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}})\xi^{\Delta+}_{\phi x}, \tag{12}$$

where $\overline{\boldsymbol{H}}$ is the detraced Hessian of the density and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ so that

$$\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}} = \kappa \frac{\cos(2\theta)}{2}, \tag{13}$$

r being the distance to the critical point and θ the angle from the eigen-direction corresponding to the first eigenvalue λ_1 of the critical point. When r goes to zero, given the properties of the ξ functions (see Fig. 5), the density trivially converges to the constraint ν.

3.1.3 *Constrained 2D spin maps*

In two dimensions, the spin is a scalar given by

$$s_z(\boldsymbol{r}) = \sum_{i,j,k} \epsilon_{ij3}\phi_{ik}x_{jk}, \tag{14}$$

where $\epsilon = \epsilon_{ij3}$ is built upon the totally anti-symmetric rank 3 Levi-Civita tensor ϵ_{ijk}. Since equation (14) is quadratic in the fields x and ϕ, equation (11) can be readily applied to compute analytically its conditional expectation. The angular momentum generated by TTT as a function of the polar position (r, θ) subject to the same critical point constraint at the origin with contrast ν, and principal

[2] see http://tinyurl.com/mmbse3z which describes an implementation in MATHEMATICA of the conditional probability.

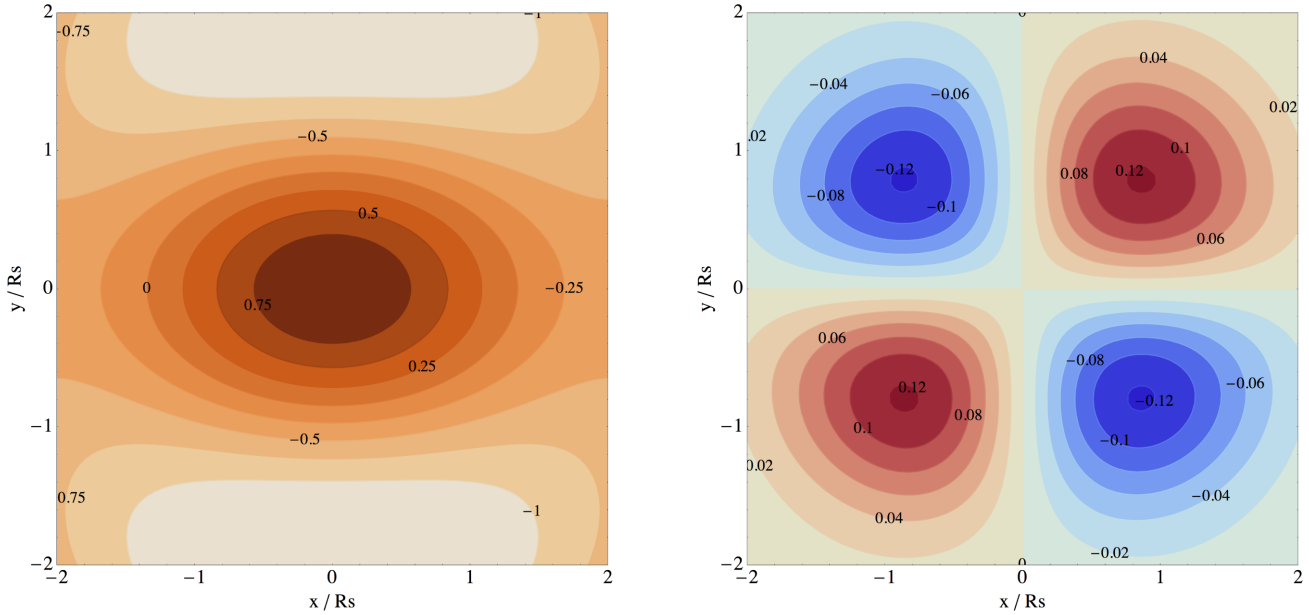


Figure 6. Left: mean density (contrast) field near a 2D peak of height $\nu = 1$, $\lambda_1 = -1$ and $\lambda_2 = -2$ for a power spectrum with index $n = 1/2$ computed from equation (12). Contours are displayed from $\delta = -1$ to 1 by step of 1/4 as labelled. The x- and y-axes are in units of the smoothing length. Right: corresponding mean spin colour coded from blue (negative) to red (positive) computed from equation (14). The flattening of the filament's cross-section induces a clear quadrupolar spin distribution in its vicinity.

curvatures (λ_1, λ_2) is given by the sum of a quadrupole ($\propto \sin 2\theta$) and an octupole ($\propto \sin 4\theta$)

$$\langle s_z|\mathrm{ext}\rangle = s_z(\boldsymbol{r}|\mathrm{ext}, \kappa, I_1, \nu),$$
$$= -16(\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \epsilon \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}}) \left(s_z^{(1)} + 2(\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}}) s_z^{(2)}\right), \quad (15)$$

where the octupolar coefficient $s_z^{(2)}$ can be written as

$$s_z^{(2)}(r) = \left(\xi_{\phi x}^{\Delta\Delta} \xi_{xx}^{\times\times} - \xi_{\phi x}^{\times\times} \xi_{xx}^{\Delta\Delta}\right),$$

and the quadrupolar coefficient $s_z^{(1)}$ reads

$$s_z^{(1)}(r) = \frac{\nu}{1-\gamma^2}\left[\left(\xi_{\phi\phi}^{\Delta+} + \gamma\xi_{\phi x}^{\Delta+}\right)\xi_{xx}^{\times\times} - \left(\xi_{\phi x}^{\Delta+} + \gamma\xi_{xx}^{\Delta+}\right)\xi_{\phi x}^{\times\times}\right]$$
$$+ \frac{I_1}{1-\gamma^2}\left[\left(\xi_{\phi x}^{\Delta+} + \gamma\xi_{\phi\phi}^{\Delta+}\right)\xi_{xx}^{\times\times} - \left(\xi_{xx}^{\Delta+} + \gamma\xi_{\phi x}^{\Delta+}\right)\xi_{\phi x}^{\times\times}\right],$$

while

$$\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \epsilon \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}} = -\kappa \frac{\sin(2\theta)}{2}. \quad (16)$$

Equation (15) is remarkably simple. As expected, the spin, s_z, is identically null if the filament is axially symmetric ($\kappa = 0$). It is zero along the principal axis of the Hessian (where $\theta = 0 \bmod \pi/2$ for which $\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \epsilon \cdot \hat{\boldsymbol{r}} = 0$). Near the peak, the anti-symmetric, $\sin(2\theta)$, component dominates, and the spin distribution is quadrupolar. For scale-invariant density power spectra with index n ($n - 4$ for the *potential*), s_z can be computed explicitly. At small separation, s_z behaves like

$$s_z \propto \kappa((n+2)\nu + \sqrt{(n+2)(n+4)}I_1)r^2 \sin(2\theta), \quad (17)$$

which shows explicitly that the quadrupolar term dominates.

Fig. 6 displays the mean density and spin map for a power-law power spectrum with index $n = 1/2$ around a 2D peak of the density field with geometry $\nu = 1$, $\lambda_1 = -1$ and $\lambda_2 = -2$.

At this stage, it is interesting to understand how much angular momentum is contained into spheres of increasing radius that would feed the forming object at different stages of its evolution. For instance, let us assume there is a small-scale overdensity at (one of the four) location of maximum angular momentum (denoted $r_\star$ hereafter) and let us filter the spin field with a top-hat window function centred on $r_\star$ and of radius R_{TH}. The resulting amount of angular momentum as a function of this top-hat scale is displayed in Fig. 7. During the first stage of evolution, the central object will acquire spin constructively until it reaches a Lagrangian size of radius $R_{\mathrm{TH}} = r_\star$ and feels the two neighbouring quadrants of opposite spin direction. The spin amplitude then decreases and becomes even negative before it is fed by the last quadrant of positive spin. The minimum is reached for radius around $2.4r_\star$. This result does not change much with the contrast and the geometry of the peak constraint. Fig. 7 is the Lagrangian counterpart of fig. 4 of Laigle et al. (2015, or fig. 7 of Pichon & Bernardeau 1999) which displays the quadrant of vorticity in the vicinity of filaments.

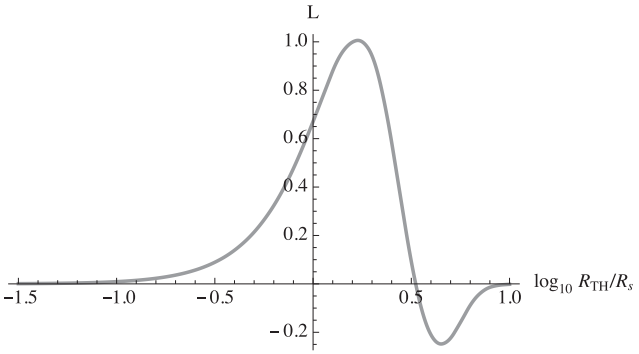


Figure 7. Evolution of the amount of algebraic angular momentum in sphere of radius R_{TH} centred on $r_\star$ The density power-spectrum index is $n = -3/2$, the height of the peak in (0, 0) is $\nu = 1$ and principal curvatures $\lambda_1 = -1, \lambda_2 = -2$. The amplitude of the spin is normalized by its maximum value around $R_{\mathrm{TH}} = r_\star$

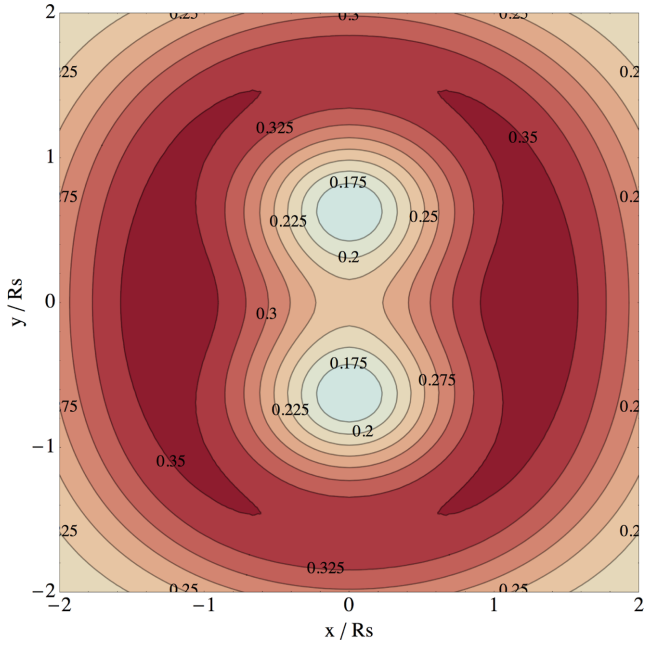


Figure 8. 2D spin dispersion (defined in equation 18) near a 2D peak of height $\nu = 1$ and curvatures $\lambda_1 = -1$ and $\lambda_2 = -2$ for a power spectrum with index $n = 1/2$

3.1.4 Cosmic variance on spin

On top of the mean spin, one can also compute the dispersion of the spin described by

$$\sigma(\boldsymbol{r}) = \sqrt{\langle s_z^2(\boldsymbol{r})\rangle - \langle s_z(\boldsymbol{r})\rangle^2}\,. \tag{18}$$

A map of this spin dispersion is shown on Fig. 8. Comparing Fig. 8 to Fig. 6, we see that spin direction fluctuates along the major axis of the filament cross-section, and best defined along its minor axis. As the conditional statistics is Gaussian, the whole spin statistics (third moments,. . .) can in principle be similarly computed.

3.1.5 Zel'dovich mapping of the Spin

Fig. 9 displays the image of the initial density field (resp. initial spin field) translated by a Zel'dovich displacement. The displacement is proportional to (ϕ_1, ϕ_2) here and its expectation given a central peak is trivially computed from the conditional PDFs. The resulting quadrupolar caustics is qualitatively similar to the quadrupolar geometry of the vorticity field measured in numerical simulations (Laigle et al. 2015). Indeed, as discussed in that paper, there is a dual relationship between such Eulerian vorticity maps and the geometry of the spin distribution within the neighbouring patch of a 3D saddle point.

3.2 Transition mass for long filaments

Up to know we assumed that the geometry of the critical point was given. Let us now build the joint statistics of the spin and the mass near 2D peaks.

3.2.1 Geometry of the most likely cross-section

Let us now study what should be the typical geometry of a peak. Following Pogosyan et al. (2009), it is straightforward to derive the PDF for a point to have height ν and geometry κ, I_1 as in their notation $J_2 = \kappa^2$ so that

$$\mathcal{P}(\nu,\kappa,I_1) = \frac{\kappa}{\pi\sqrt{1-\gamma^2}}\exp\left(-\frac{1}{2}\left(\frac{\nu+\gamma I_1}{\sqrt{1-\gamma^2}}\right)^2 - \frac{1}{2}I_1^2 - \kappa^2\right).$$

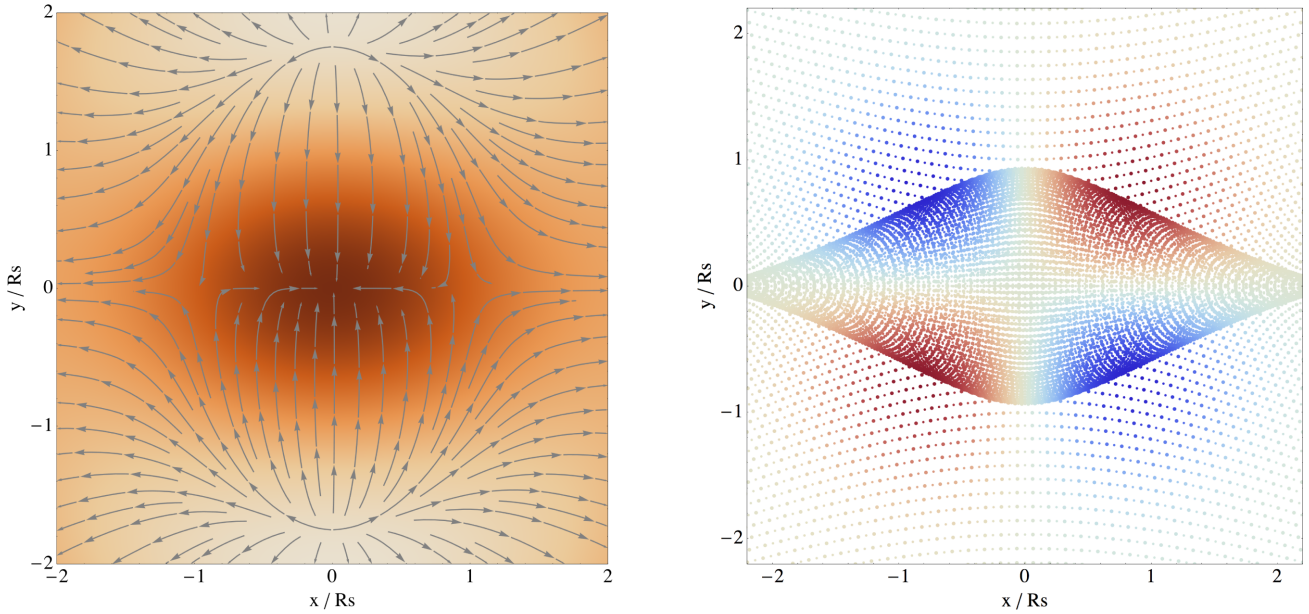


Figure 9. Left: stream lines of the 2D velocity field (defined as the potential gradient) near a 2D peak of height $\nu = 1$ and curvatures $\lambda_1 = -1$ and $\lambda_2 = -2$ for a power spectrum with index $n = 1/2$. Right: Zel'dovich mapping of the spin distribution. There is a good qualitative agreement between the vorticity section presented in Laigle et al. (2015) and this spin map.

Now the PDF for a *peak* to have height ν and geometry κ, I_1 becomes:

$$\mathcal{P}(\nu,\kappa,I_1|pk) = \frac{\sqrt{3}\kappa|(I_1-\kappa)(I_1+\kappa)|}{2\pi\sqrt{1-\gamma^2}}\Theta(-\kappa-I_1) \times \exp\left(-\frac{1}{2}\left(\frac{\nu+\gamma I_1}{\sqrt{1-\gamma^2}}\right)^2-\frac{1}{2}I_1^2-\kappa^2\right). \quad (19)$$

The maximum of this PDF is trivially reached for $\bar{\nu}=\sqrt{7/3}\,\gamma$, $\bar{\kappa}=\sqrt{1/3}$ and $\bar{I}_1=-\sqrt{7/3}$.

3.2.2 The size and area of constant polarity quadrants

From equation (15), it appears clearly that the extension of the region of influence of the critical point is limited, and peaks within each quadrant at some specific $(r_\star,\theta_\star)$ position. Moreover, for small enough κ, the quadrupole dominates, and the extremum is along $\theta=\pi/4$. It is therefore possible to use $r_\star$ to define an area in which the spin is significantly non-zero within each quadrant. Let us compute $r_\star$, as the radius for which $s_z(\theta=\pi/4)$ is maximal as a function of r.[3] The area of a typical quadrant, in which the spin has the same orientation, can then simply be expressed as

$$\mathcal{A}=\pi r_\star^2\,, \quad (20)$$

where $r_\star=r_\star(\nu,\kappa)$ is the position of a maximum of angular momentum from the peak. Because of the quadrupolar anti-symmetric geometry of the angular momentum distribution near the saddle point, it is typically twice as small (in units of the smoothing length) as one would naively expect.

For power-law density power spectrum with spectral index in the range $n\in\,]-2,2]$, a good fit to its scaling is given by

$$\frac{r_\star}{R_s}\approx\frac{3}{250}(n-5)^2+\frac{13}{10}\,, \quad (21)$$

where $r_\star$ was computed for the mean geometry given by $\bar{\nu}=\sqrt{7/3}\,\gamma$, $\bar{\kappa}=\sqrt{1/3}$ and $\bar{I}_1=-\sqrt{7/3}$.

3.2.3 Critical mass scaling

The critical mass is the mass of maximum spin alignment. In simulation, it has been shown by Laigle et al. (2015) to be $M_{\rm crit}\approx 10^{12}\mathrm{M}_\odot$ at redshift 0. The authors claimed that the critical mass is related to the mass contained in a typical quadrant of vorticity. In this work, we have computed in Lagrangian space the typical area of a quadrant (see equation 20). This area is a function of the smoothing scale. In order to compute it, we need to define a scale. It is reasonable that the maximum spin alignment should be reached for filament that has just collapsed at redshift 0. Indeed, for larger scale filaments, part of the haloes do not lie inside the filament but in the nearby wall which will therefore decrease the mean spin-filament alignment. In previous sections, we focused on $\nu=0.9$ filaments. The model of the cylindrical collapse then say that those filaments have just collapsed at redshift 0 for a top-hat initial smoothing scale $\sigma(R_{\rm TH})=1.6$ which corresponds to a smoothing length $R_{\rm TH}=2.2$ Mpc h^{-1}. We can therefore compute the corresponding $r_\star$ which is $r_\star\approx 1.6R_s\approx 0.7R_{\rm TH}=1.5$ Mpc h^{-1} and corresponds to the mass $M_{\rm crit}=\frac{4}{3}\pi r_\star^3\rho_c\Omega_m\approx 1.5\,10^{12}\,\mathrm{M}_\odot$, in good agreement with the value measured in simulations. Its redshift evolution is also predicted by the formalism through the cylindrical collapse and could be compared to simulation in future works.

Note that this line of reasoning could be made more rigorous by adding new ingredients in the formalism: a peak constraint at the location of the spin with a smoothing length $R_h<R_s$ so that one can vary R_h (without any assumption on the additivity property of the spin) and see how the spin changes. This formalism can be implemented in two dimensions (see Appendices A2 and A3) and leads to the same order of magnitude for $r_\star$.

[3] Setting $\theta=\pi/4$ effectively neglect the octupolar part of s_z.

4 3D SPIN NEAR AND ALONG FILAMENTS

Let us now turn to the truly three dimensional theory of tidal torques in the vicinity of a typical filament-saddle point. Beyond the obvious increased realism, the main motivation is that the 3D saddle theory fully captures the mass transition.

In three dimensions, we must consider two competing processes. If we vary the radius corresponding to the Lagrangian patch centred on the running point, we have a spin-up (along $\boldsymbol{e}_z$) arising from the running to wall running to saddle tidal misalignment and a second spin-up (along $\boldsymbol{e}_\phi$) arising from running to filament-running peak tidal misalignment. To each position in the vicinity of the central saddle point, we can assign $M(\boldsymbol{r})$ together with $\cos\mu_z(\boldsymbol{r})$ and $\cos\mu_\phi(\boldsymbol{r})$, the cosines of the angle between the spin of the patch and the $\boldsymbol{e}_z$ and $\boldsymbol{e}_\phi$ direction, respectively. Eliminating $\boldsymbol{r}$ yields $\cos\mu_z(M)$ and $\cos\mu_\phi(M)$ and therefore yields an estimate of the transition mass.

4.1 Spin distribution along and near filaments

The formalism developed in Section 3 can easily be extended to three dimensions. A critical (saddle) point constraint is now imposed. This critical point is defined by its geometry, namely its height ν and eigenvalues $\lambda_1\geq\lambda_2\geq\lambda_3$. Note that such a critical point is a filament-type saddle point if $\lambda_1\geq 0\geq\lambda_2\geq\lambda_3$. In what follows, we decouple the trace from the detraced part of the density Hessian and therefore define the three curvature parameters $I_1=\lambda_1+\lambda_2+\lambda_3$, $\kappa_1=\lambda_1-\lambda_2$ and $\kappa_2=\lambda_2-\lambda_3$.

4.1.1 Mean density field around a critical point

The resulting mean density (contrast) field subject to that critical point constraint becomes (in units of σ_2):

$$\delta(\boldsymbol{r}|\mathrm{crit},I_1,\kappa_1,\kappa_2,\nu)=\frac{I_1\left(\xi_{\phi x}^{\Delta\Delta}+\gamma\xi_{\phi\phi}^{\Delta\Delta}\right)}{1-\gamma^2}+\frac{\nu\left(\xi_{\phi\phi}^{\Delta\Delta}+\gamma\xi_{\phi x}^{\Delta\Delta}\right)}{1-\gamma^2} + \frac{15}{2}\left(\hat{\boldsymbol{r}}^{\mathrm{T}}\cdot\overline{\boldsymbol{H}}\cdot\hat{\boldsymbol{r}}\right)\xi_{\phi x}^{\Delta+}\,, \quad (22)$$

where again $\overline{\boldsymbol{H}}$ is the *detraced* Hessian of the density and $\hat{\boldsymbol{r}}=\boldsymbol{r}/r$ and we define in 3D $\xi_{\phi x}^{\Delta+}$ as

$$\xi_{\phi x}^{\Delta+}=\langle\Delta x\,\phi^+\rangle, \quad (23)$$

with $\phi^+=\phi_{11}-(\phi_{22}+\phi_{33})/2$. The other ξ functions are defined in the same way as in two dimensions (see equations 5) and displayed on Fig. 10. Note also that $\hat{\boldsymbol{r}}^{\mathrm{T}}\cdot\overline{\boldsymbol{H}}\cdot\hat{\boldsymbol{r}}$ is a scalar defined explicitly as $\sum_{ij}\hat{r}_i\overline{H}_{ij}\hat{r}_j$. Fig. 11 displays the mean density field around a typical filament-type saddle point. The elongation of the filament along the Oz-axis together with the flattening of the filament in the plane of the wall (Oxz) are clearly visible on this figure.

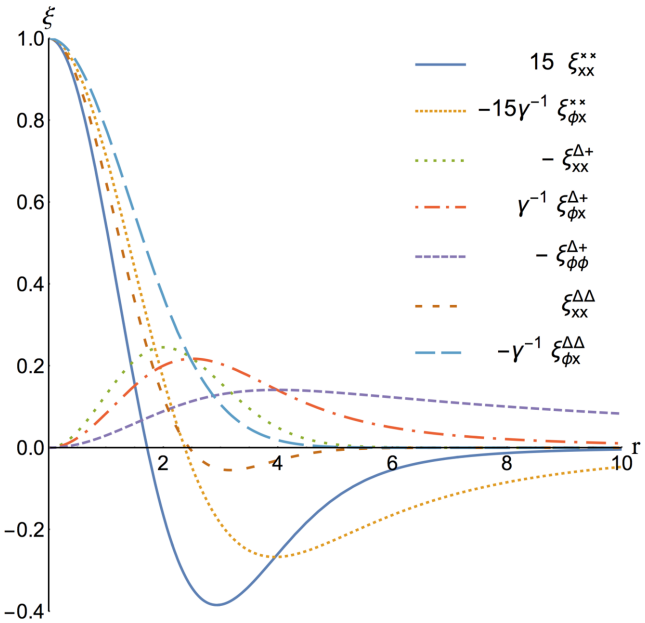


Figure 10. Two-point correlation functions as a function of the separation r in units of the smoothing length for a power-law 3D power spectrum with spectral index $n = -2$, i.e. $\gamma = \sqrt{3}/3$. As in two dimensions, the correlations are rescaled at the origin. See also Fig. C1.

4.1.2 *Mean spin field around a critical point*

As in two dimensions, the expected spin can also be computed. In three dimensions, the spin, $\boldsymbol{s}$, is a vector which components are given by

$$s_i = \sum_{j,k,l} \epsilon_{ijk} x_{kl} \phi_{lj} \,, \tag{24}$$

with ϵ the rank 3 Levi-Civita tensor. It is found to be orthogonal to the separation and can be written as the sum of two terms

$$\boldsymbol{s}(\boldsymbol{r}|\mathrm{crit}, I_1, \kappa_1, \kappa_2, \nu) = -15(\boldsymbol{s}^{(1)} + \boldsymbol{s}^{(2)}) \cdot (\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \epsilon \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}}) \,, \tag{25}$$

where $\boldsymbol{s}^{(1)}$ is a scalar operator that depends on the height ν and trace of the Hessian I_1

$$\boldsymbol{s}^{(1)} = \left(\frac{\nu}{1-\gamma^2} \left[\left(\xi^{\Delta+}_{\phi\phi} + \gamma \xi^{\Delta+}_{\phi x} \right) \xi^{\times\times}_{xx} - \left(\xi^{\Delta+}_{\phi x} + \gamma \xi^{\Delta+}_{xx} \right) \xi^{\times\times}_{\phi x} \right] \right.$$
$$\left. + \frac{I_1}{1-\gamma^2} \left[\left(\xi^{\Delta+}_{\phi x} + \gamma \xi^{\Delta+}_{\phi\phi} \right) \xi^{\times\times}_{xx} - \left(\xi^{\Delta+}_{xx} + \gamma \xi^{\Delta+}_{\phi x} \right) \xi^{\times\times}_{\phi x} \right] \right) \mathbb{I}_3 ,$$

and $\mathrm{s}^{(2)}$ a combination of a matricial and a scalar operator that depends on the detraced part of the Hessian

$$\boldsymbol{s}^{(2)} = -\frac{5}{8} \Big[2 \left(\left(\xi^{\Delta+}_{\phi x} - \xi^{\Delta\Delta}_{\phi x} \right) \xi^{\times\times}_{xx} - \left(\xi^{\Delta+}_{xx} - \xi^{\Delta\Delta}_{xx} \right) \xi^{\times\times}_{\phi x} \right) \overline{\boldsymbol{H}}$$
$$+ \left(\left(7\xi^{\Delta\Delta}_{xx} + 5\xi^{\Delta+}_{xx} \right) \xi^{\times\times}_{\phi x} - \left(7\xi^{\Delta\Delta}_{\phi x} + 5\xi^{\Delta+}_{\phi x} \right) \xi^{\times\times}_{xx} \right)$$
$$\times (\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}}) \mathbb{I}_3 \Big]$$

with $\mathbb{I}_3$ the identity matrix, operating on the vector

$$\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \epsilon \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}} = \sum_{ikl} \hat{\boldsymbol{r}}_i \epsilon_{ijk} \overline{\boldsymbol{H}}_{kl} \hat{\boldsymbol{r}}_l \,. \tag{26}$$

Note that the dependence with the distance r is encoded in the two-point correlation functions, ξ, while the geometry of the critical point is encoded in the terms corresponding to the peak height, trace and detraced part of the Hessian and the orientation of the separation is in $\hat{\boldsymbol{r}}$. Equation (25) is also remarkably simple: as expected the symmetry of the model induces zero spin along the principal directions of the Hessian (where $\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \epsilon \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}} = 0$) and a point reflection symmetry ($\hat{\boldsymbol{r}} \to -\hat{\boldsymbol{r}}$). Note that the correlation functions, ξ can be evaluated for arbitrary power spectra (such as power laws, see Appendix C2, or Λ cold dark matter (ΛCDM), see Appendix C3), hence equation (25) is completely general.

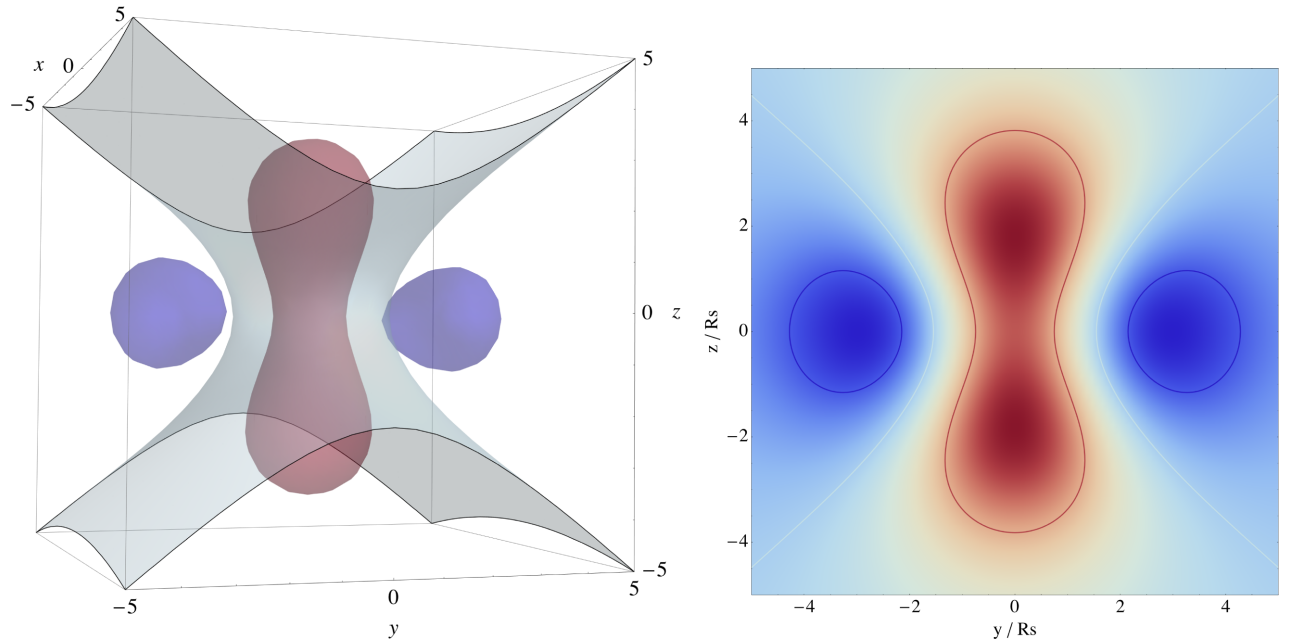


Figure 11. Mean density around a filament-saddle point of height $\nu = 1.25\gamma$, $\lambda_1 = 0.31, \lambda_2 = -0.56, \lambda_3 = -1$ for a power-law 3D power spectrum with spectral index $n = -2$, i.e. $\gamma = \sqrt{3}/3$. The Ox-, Oy- and Oz-axes are in units of the smoothing length and z is the direction of the filament while the wall is in the plane Oxz. Contours represent the isocontours of the density $\delta = -0.3, 0, 0.5$ from blue to red. A 3D view is displayed on the left-hand panel and a cut in the plane Oyz is shown on the right-hand panel.

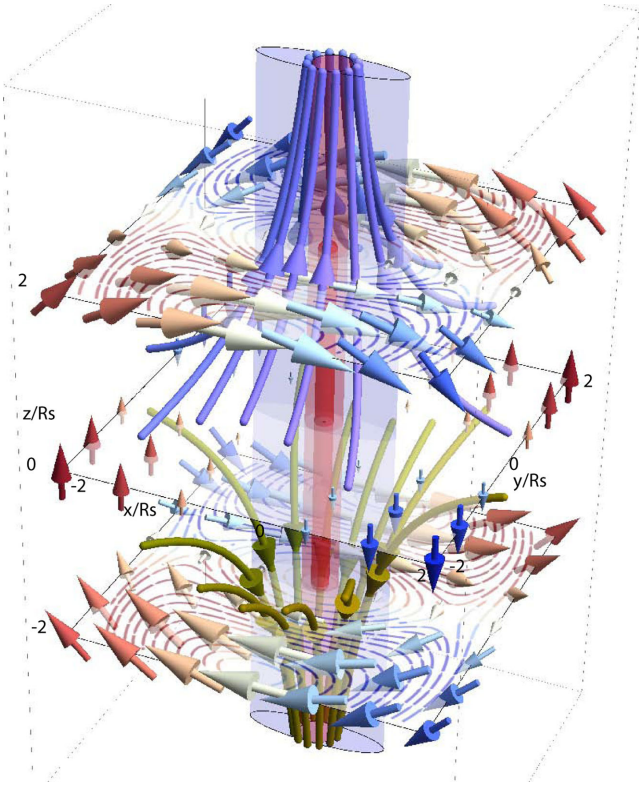


Figure 12. The velocity and spin flow near a vertical filament (in red) embedded in a (purple) wall for $(x, y, z) \in [-2R_s, 2R_s]^3$. The purple and green flow lines trace the (Lagrangian) 3D velocities (upwards and downwards, respectively). The red and blue arrows show the spin 3D distribution, while the three horizontal cross-sections show spin flow lines in the corresponding plane. Note that the spin is along $\boldsymbol{e}_z$ in the mid-plane and along $\boldsymbol{e}_\phi$ away from it, and that it rotates in opposite direction above and below the mid-plane. See also the interactive version at http://www.iap.fr/users/pichon/AM-near-saddle.html.

For scale-invariant density power spectra with index n ($n - 4$ for the *potential*), $\boldsymbol{s}$ can be computed explicitly. At small separation, the term proportional to $\hat{\boldsymbol{r}}^{\mathrm{T}} \cdot \overline{\boldsymbol{H}} \cdot \hat{\boldsymbol{r}}$ goes like r^4 and is thus negligible compare to the rest (that scales like r^2). The spin coordinates in the frame of the Hessian are therefore quadrupolar

$$\boldsymbol{s} \propto (f(\lambda_i, \nu, n)yz, g(\lambda_i, \nu, n)xz, h(\lambda_i, \nu, n)xy)\,. \tag{27}$$

Fig. 12 illustrates the mean spin geometry around a typical saddle point. All the symmetry properties (anti-symmetry, octopole,...) described in this section are clearly seen on this figure. In the plane of the saddle point, spins are aligned with the filament direction. When moving towards the nodes, the spins become more and more perpendicular (and more and more along $\boldsymbol{e}_\phi$).

4.1.3 Cosmic variance on 3D spin

It is of interest to also study the variance of the spin alignment $\sigma(\boldsymbol{r}|\mathrm{ext}, I_1, \kappa_1, \kappa_2, \nu)$ defined as

$$\sigma = \sqrt{\langle \cos^2\theta \rangle - \langle \cos\theta \rangle^2}\,, \tag{28}$$

where $\cos\theta = \boldsymbol{s} \cdot \boldsymbol{e}_z / ||\boldsymbol{s}||$. It requires the numerical evaluation of a 12D integral. In contrast, the mean of the spin $\boldsymbol{s}$ (as computed in Section 4.1.2) or its square $\boldsymbol{s}^2$ can be analytically computed.

We therefore propose to approximate the dispersion of the spin alignment with the following related estimator

$$\tilde{\sigma} = \sqrt{\frac{\langle s_z^2 \rangle - \langle s_z \rangle^2}{\langle \boldsymbol{s} \cdot \boldsymbol{s} \rangle}}\,, \tag{29}$$

where s_z is the component of the spin along the z-axis, i.e. along the filament direction. For the sake of readability, we do not write down the result of the integration here but display in Fig. 13 the map of the alignment dispersion $\tilde{\sigma}$ around a typical saddle point. This standard deviation is roughly constant around $\approx$0.6 and decreases to $\approx$0.3 in the close vicinity of the saddle point. Note that the spin direction is again best defined along its minor axis. This would be the best place to measure spin alignments in observations.

4.2 Mean saddle-point geometry

Here, we want to compute the mean values of ν, $\lambda_1 < \lambda_2 < 0 < \lambda_3$ of a typical saddle point of filament type. Let us start from the so-called Doroshkevich formula for the PDF of these variables:

$$\mathcal{P}(\nu, \lambda_i) = \frac{135\,(5/2\pi)^{3/2}}{4\sqrt{1-\gamma^2}} \exp\left[-\frac{1}{2}\zeta^2 - 3I_1^2 + \frac{15}{2}I_2\right] \times (\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)(\lambda_2 - \lambda_1)\,,$$

where $\zeta = (\nu + \gamma I_1)/\sqrt{1-\gamma^2}$, $I_1 = \lambda_1 + \lambda_2 + \lambda_3$, $I_2 = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3$ and $I_3 = \lambda_1\lambda_2\lambda_3$. Subject to a saddle-point constraint, this PDF becomes

$$\mathcal{P}(\nu, \lambda_i|\mathrm{sad}) = \frac{540\sqrt{5\pi}\mathcal{P}(\nu, \lambda_i)}{29\sqrt{2} + 12\sqrt{3}} I_3 \Theta(\lambda_3)\Theta(-\lambda_2)\,, \tag{30}$$

after imposing the condition of saddle point $|\det \partial_i \partial_j \delta| \delta_D(\nabla\delta)\Theta(\lambda_3)\Theta(-\lambda_2)$ for which as the gradient is decoupled from the density and the Hessian, only the condition on the sign of the eigenvalues and the determinant contribute. From this PDF, it is straightforward to compute the expected value of the density and the eigenvalues at a saddle-point position: $\langle\nu\rangle \approx 0.76\gamma$, $\langle\lambda_1\rangle \approx -0.87$, $\langle\lambda_2\rangle \approx -0.40$ and $\langle\lambda_3\rangle \approx 0.51$. However, this saddle point does not belong to the skeleton of the density field but to its inter-skeleton (see Pogosyan et al. 2009). We thus want to impose an additional constraint which is $\lambda_2 + \lambda_3 < 0$. Let us call those saddle points 'skeleton saddles'. The PDF at those points becomes

$$\mathcal{P}(\nu, \lambda_i|\mathrm{skl}) = \frac{26460\sqrt{5\pi}\mathcal{P}(\nu, \lambda_i) I_3 \Theta(\lambda_3)}{1421\sqrt{2} - 735\sqrt{3} + 66\sqrt{42}} \Theta(-\lambda_2 - \lambda_3)\,. \tag{31}$$

The expected value of the density and the eigenvalues at a skeleton saddle position now becomes $\langle\nu\rangle \approx 1.25\gamma$, $\langle\lambda_1\rangle \approx -1.0$, $\langle\lambda_2\rangle \approx -0.56$ and $\langle\lambda_3\rangle \approx 0.31$.

4.3 Spin flip: from spatial to mass transition

The geometry of the spin distribution near a typical skeleton saddle point (as defined by equation (31)) allows us to compute the mean alignment angle between the spin and the filament (see Section 4.3.1 below). In turn, the shape of the density profile in the vicinity of the same critical point, together with an extension of the Press–Schechter theory involving a filament-background split, allows us to estimate the 'typical' mass of the DM haloes forming in any spatial position around the saddle point (Section 4.3.2 below). The alignment-angle map and the typical-mass map will together yield a prediction for the transition mass.

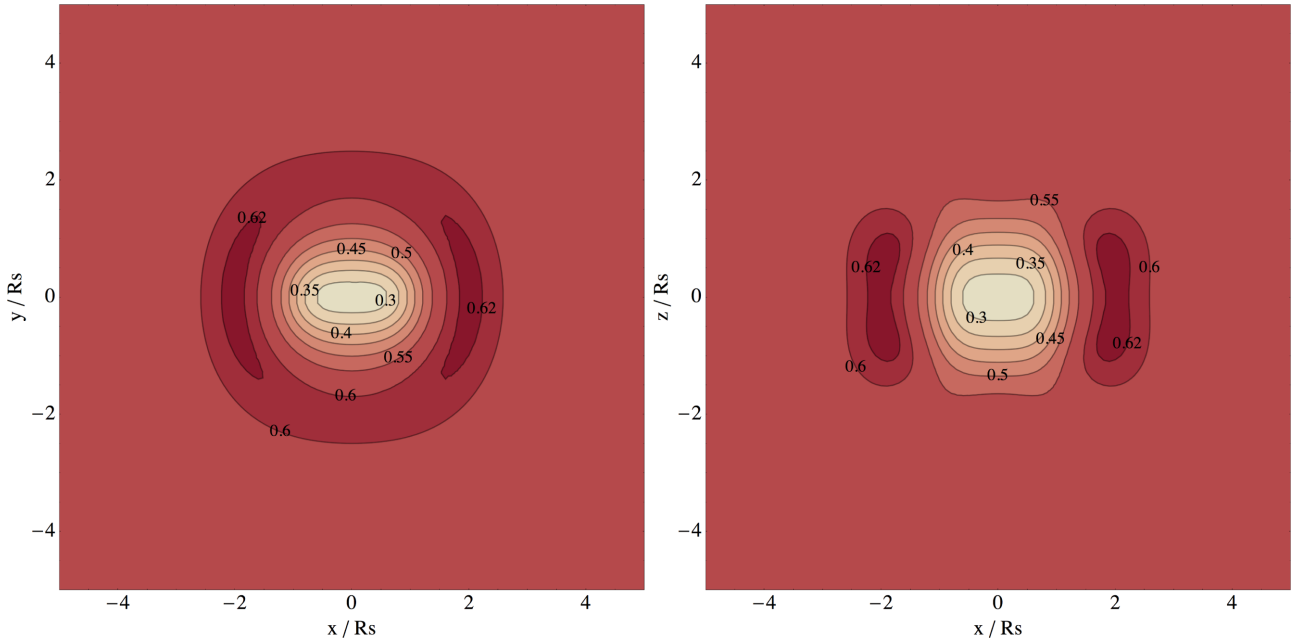


Figure 13. Alignment dispersion $\tilde{\sigma}$ (defined in equation 29) around a typical filament-type saddle point of height $\nu = 1.25\gamma, \lambda_1 = 0.31, \lambda_2 = -0.56, \lambda_3 = -1$ for a power-law 3D power spectrum with spectral index $n = -2$. The left-hand panel displays a cut along the plane, *Oxy*, of the saddle point and right-hand panel along the plane, *Oxz*, of the wall. The uncertainty on spin direction is smallest near the saddle.

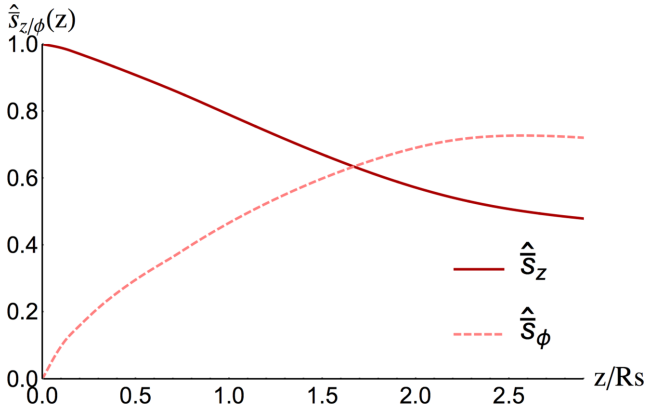


Figure 14. measure the height z corresponding to the transition from aligned to perpendicular to the filaments. The amplitude in both direction is averaged by plane following equation (32). The transition curvilinear coordinate is $z_{\rm tr} = 1.5R_{\rm s}$.

4.3.1 Spin flip along filaments

Section 4.1 showed that the mean spin flips from alignment in the plane of the saddle point to orthogonality when going towards the nodes. This can be quantified by measuring the curvilinear coordinate along the filament at which the spin flips.

Let us consider the mean *modulus* of the projection of the spin along the $\boldsymbol{e}_z$- and $\boldsymbol{e}_\phi$-axes within a plane of height z

$$\hat{\bar{s}}_{z/\phi}(z) = \int \mathrm{d}x\mathrm{d}y \left|\bar{s}_{z/\phi}(\boldsymbol{r})\right| / ||\boldsymbol{s}(\boldsymbol{r})|| \,. \tag{32}$$

Fig. 14 displays $\hat{\bar{s}}_{z/\phi}(z)$ as a function of z along the filament. Let us define $\hat{\theta}$ the flip angle so that

$$\cos\hat{\theta}(z) = \frac{\hat{\bar{s}}_z(z)}{\sqrt{\hat{\bar{s}}_z(z)^2 + \hat{\bar{s}}_\phi(z)^2}} = \frac{1}{\sqrt{2}} \,. \tag{33}$$

In Fig. 14, this flip angle is found to occurs around $z = 1.5R_{\rm s}$ which is very close to the $r_\star$ measured in two dimensions (see Section 3.2).

Alternatively, one can also compute at each position the mean alignment with the filament direction $\boldsymbol{e}_z$

$$\cos\theta(\boldsymbol{r}) = \frac{\boldsymbol{s}(\boldsymbol{r}|\mathrm{crit}) \cdot \boldsymbol{e}_z}{||\boldsymbol{s}(\boldsymbol{r}|\mathrm{crit})||} \,. \tag{34}$$

The result is shown on the right-hand panel of Fig. 15. Spins tend to align with the filament (region in red) in the plane of the saddle point and becomes perpendicular to it when moving towards the nodes (region in blue). This is a transition in Lagrangian space. Section 4.3.2 shows how to convert it into a transition in mass.

4.3.2 Halo mass gradient along filament

The local mass distribution of haloes is expected to vary along the large-scale filament due to changes in the underlying long-wave density. In the linear regime, the typical overdensity near the end points (nodes) of the filament, where it joins the protocluster regions, may exceed the typical overdensity near the saddle point by a factor of 2 (Pogosyan et al. 1998). During epochs before the whole filamentary structure has collapsed, this leads to a shift in the hierarchy of the forming haloes towards larger masses near the filament end points (the clusters) relative to the filament middle point (the saddle). This can be easily understood using the formalism of barrier crossing (Peacock & Heavens 1990; Bond et al. 1991; Paranjape, Lam & Sheth 2012; Musso & Sheth 2012), which associates the density of objects of a given mass to the statistics for the random walk of halo density as the field is smoothed with decreasing filter sizes. Specifically, these authors predict the first upcrossing probability for the critical threshold at the filter scale corresponding to the mass of interest. The precise outcome of the formalism depends on the spectral properties of the field and the form of the smoothing filter, however it is clear that, in general, decreasing the barrier threshold increases the probability that such first upcrossing

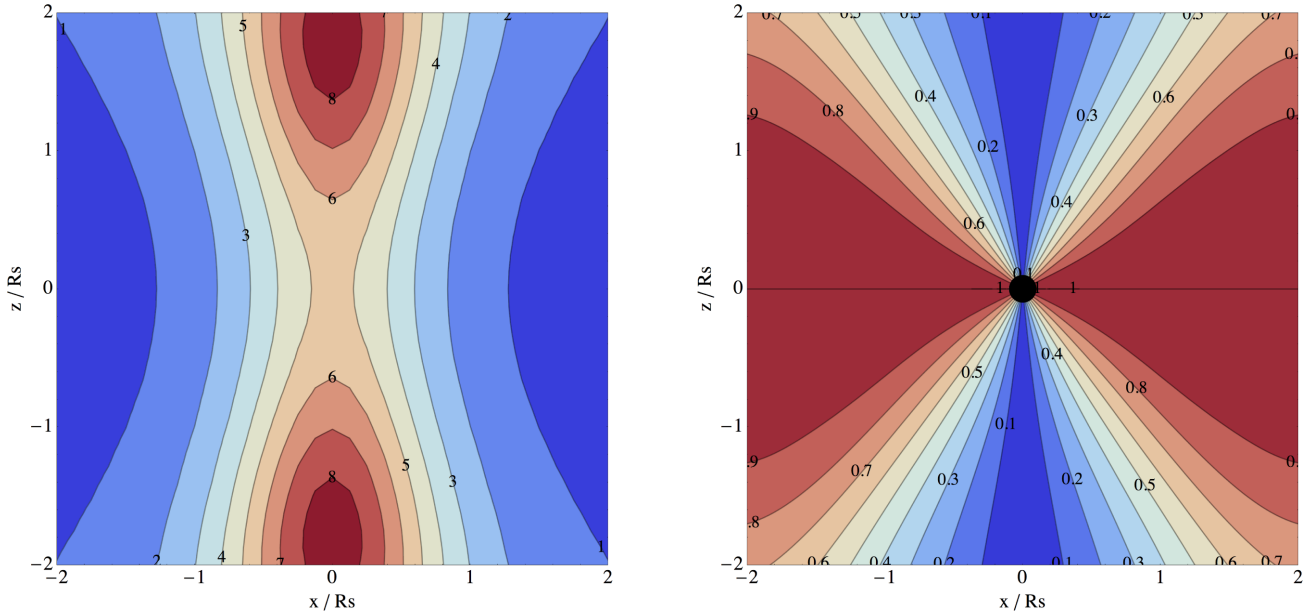


Figure 15. Left: cross-section of $M_{\rm p}(r, z)$ (in units of $10^{12}\,{\rm M}_{\odot}$) along the most likely filament and in the direction $x = y$. Right: corresponding cross-section of $\langle \cos\hat{\theta} \rangle(r, z)$, the normalized component of the spin aligned with the filament. The black dot represents the position of the saddle point. The mass of haloes increases towards the nodes, while the spin flips.

will happen at large smoothings, i.e. large mass. A larger fraction of the Lagrangian space will then belong to large-mass haloes, at the expense of the low-mass ones.

Following the presentation of Paranjape et al. (2012) of the Peacock–Heavens (Peacock & Heavens 1990) approximation – that was found to fit numerical simulations rather well, the number density of dark haloes in the interval $[M, M + {\rm d}M]$ is

$$\frac{{\rm d}n(M)}{{\rm d}M}{\rm d}M = \frac{\rho}{M} f(\sigma^2, \delta_{\rm c}){\rm d}\ln\sigma^2\,, \tag{35}$$

where $f(\sigma^2, \delta_{\rm c})$ is given by the function

$$f(\sigma^2, \delta_{\rm c}) = \exp\left(\frac{1}{\Gamma}\int_0^{\sigma^2}\frac{{\rm d}s'}{s'}\ln p(s', \delta_{\rm c})\right) \times \left(-\sigma^2\frac{{\rm d}p(\sigma^2, \delta_{\rm c})}{{\rm d}\sigma^2} - \frac{1}{\Gamma}p(\sigma^2, \delta_{\rm c})\ln p(\sigma^2, \delta_{\rm c})\right). \tag{36}$$

Here, σ^2 is the variance of the density fluctuations smoothed at the scale corresponding to M and $p(\sigma^2, \delta_{\rm c}) \equiv 1/2(1 + {\rm erf}(\delta_{\rm c}/\sqrt{2}\sigma))$ is the probability of a Gaussian process with variance σ^2 to yield value below some critical threshold $\delta_{\rm c}$. In equation (36), Γ is the parameter dependent on the filtering scale and, to less extend the underlying power spectrum, that specifies how correlated the density values at the same point when smoothed at different scales are. For Gaussian filter, the value $\Gamma \approx 4$ is advocated.

The overall mass distribution of haloes is well described by the choice $\delta_{\rm c} = 3/5\,(3\pi/2)^{2/3} = 1.681$, motivated by the so-called spherical collapse model. When haloes form on top of a large-scale structure background, however, the long-wave overdensity $\bar{\delta}(z)$ adds to the overdensity in the protohalo peaks. The effect on halo mass distribution, in this so-called peak-background-split approach, can be approximated as a *shifted* threshold $\delta_{\rm c}(z) = 1.681 - \bar{\delta}(z)$ for halo formation. In Fig. 16, we show that, as expected, the result of this long-wavelength mode is a shift of the halo mass distribution towards larger masses. This shift can be characterized by the dependence on the threshold of $M_*(\delta_{\rm c})$, defined as $\sigma_*(M_*) = \delta_{\rm c}$, or of the mass $M_{\rm p}(\delta_{\rm c})$ that corresponds to the peak of $f(\sigma^2, \delta_{\rm c})$, i.e. the variance $\sigma_{\rm p}^2(z)$ defined by

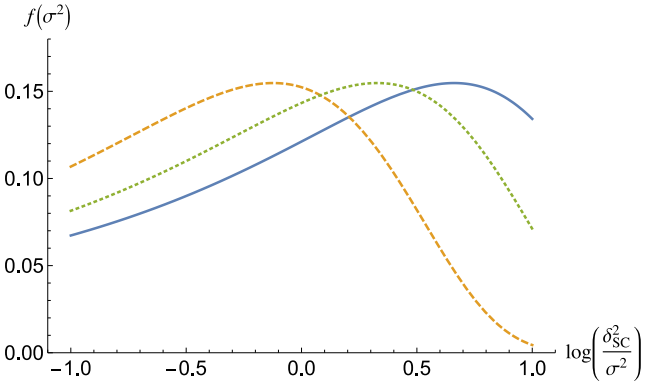


Figure 16. Mass distribution for three values of density threshold, $\delta_{\rm c} = 1.681$, 1, 0.681 from left (yellow dashed line) to right (blue solid line). The displayed function $f(\sigma^2)$ is defined in equation (36).

$$\sigma_{\rm p}^2(z) \equiv \underset{\sigma^2}{\rm argmax}(f(\sigma^2, \delta_{\rm c}(z)))\,. \tag{37}$$

Fig. 17, right-hand axis, shows these two characteristic variances as functions of the threshold, $\delta_{\rm c}$.

The link to cosmology is established by relating the variance σ^2 to the mass of the objects. If a background field is absent, the variance is just the integral of the power spectrum $P(k)$ smoothed over a sphere of mass M

$$\sigma^2 = \sigma^2(M, Z) \equiv D^2(Z)\int k^2{\rm d}k\,P(k)W_{\rm TH}^2\left((3\,M/4\pi\bar{\rho})^{1/3}\right),$$

where $D(Z)$ is the linear growing mode of perturbations as a function of redshift Z and $W_{\rm TH}$ is the top-hat filter. However, when large-scale structures are considered as fixed background, the variance of the

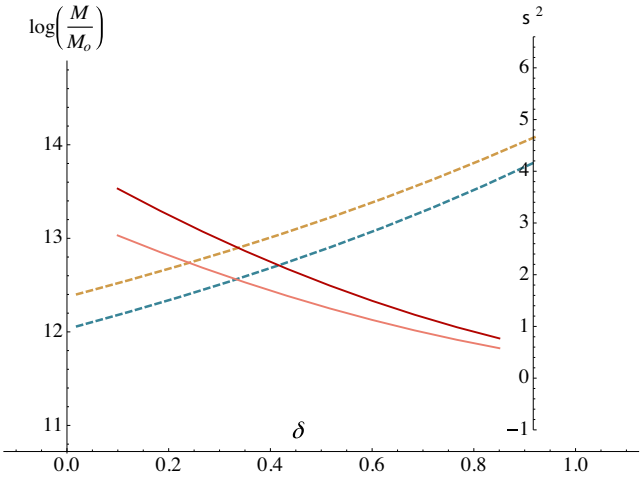


Figure 17. Characteristic variances (plain) and M_* (yellow dashed) and M_p halo masses (blue dashed) as functions of the large-scale density in the peak-background-split picture.

relevant small-scale density fluctuations that are responsible for object formation is reduced, approximately as

$$\sigma^2 \approx \sigma^2(M) - \sigma^2(M_\mathrm{LSS})\,, \tag{38}$$

where $\sigma^2(M_\mathrm{LSS})$, given as well as $\sigma^2(M)$ by equation (38), is the unconstrained variance at the scales at which we have defined the background large-scale density. This correction is negligible when there is distinct scale separation between non-linear forming objects and the large-scale density, i.e. $\bar{\delta}(x) \ll 1.681$ but becomes important, truncating the mass hierarchy at M_LSS, whenever large-scale structures are themselves non-linear.

On Fig. 17, left-hand axis, the variances are converted into masses, $M_\star$ and M_p according to equation (38). We choose here $\sigma_8 = 0.8$, $Z = 0$, we define the mass in a $8h^{-1}$ Mpc comoving sphere for the best-fitting cosmological mass density and we approximate the spectrum with a power law of index $n = -2$, which allows to solve equation (38) explicitly, giving the $M(\sigma)$ relation as

$$M(\sigma, Z) = 2.6 \times 10^{14}\,\mathrm{M}_\odot \left(\frac{\sigma^2 + \sigma^2(M_\mathrm{LSS})}{\sigma_8^2 D(Z)^2}\right)^{-\frac{3}{n+3}}. \tag{39}$$

We consider filaments to be defined with $R = 5h^{-1}$ Mpc Gaussian smoothing, which gives $\sigma^2(M_\mathrm{LSS}) \approx 0.66$. The evolution of $M_\mathrm{p}(\bar{\delta}, Z)$ follows from putting equation (37) into equation (39).

4.3.3 Spin orientation versus mass

From the above described M_p–δ relation, one can attribute a mass to each position depending on the value of the mean density at that location. The result is illustrated in Fig. 15 where the left- and right-hand panels display, respectively, the mass map and the spin alignment map around a typical saddle point. Eliminating the spatial position, $\boldsymbol{r}$, between these two maps yields $\langle \cos\theta \rangle$ as a function of M_p as shown on Fig. 18. The transition mass, $\mathcal{M}_\mathrm{tr}$ for spin flip ($\langle \cos\theta \rangle = 0.5$) is found to be of the order of $4 \times 10^{12}\,\mathrm{M}_\odot$, assuming a smoothing scale of 5 Mpc h^{-1}, as used in Codis et al. (2012). This mass is in qualitative agreement with the transition mass found in that paper, all the more so as the redshift evolution of this transition mass will also be consistent (scaling as the mass of non-linearity).

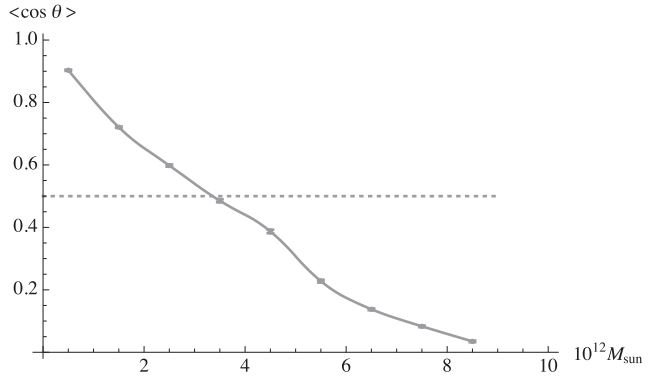


Figure 18. Mean alignment as a function of mass for a smoothing scale for filaments of 5 Mpc h^{-1}. Error bars represent the error on the mean cosine in each bin of mass for the region $[-2R_\mathrm{s}, 2R_\mathrm{s}] \times [-2R_\mathrm{s}, 2R_\mathrm{s}] \times [-2R_\mathrm{s}, 2R_\mathrm{s}]$ around a typical filament-saddle point. The flip transition mass corresponds to $4 \times 10^{12}\,\mathrm{M}_\odot$.

It is quite striking that the geometry of the saddle point alone allows us to predict this mass. The two main ingredients for success are the point reflection symmetry of the spin distribution near the most likely filament-like saddle point on the one hand, and the peak-background-split mass distribution gradient along the filament towards the nodes of the cosmic web on the other hand.

5 STATISTICS

Up to now, we have considered the neighbourhood of a *given unique* typical saddle point as a proxy for the behaviour within a GRF. In view of our finding let us now first analyse the statistics of alignment for GRF, and then for fields corresponding to their simulated cosmic evolution down to redshift zero.

5.1 Validation on GRF

Let us consider the following experiment. Let us generate 2D or 3D realizations of GRF smoothed on two successive scales, L_h and $L_\mathrm{s} \gg L_\mathrm{h}$. In the first maps, let us build a catalogue of positions, $\boldsymbol{r}_\mathrm{h}$ and heights, ν_h corresponding to 'small-scale' peaks. From the second maps, let us identify the loci, $\boldsymbol{r}_\mathrm{s}$ of the corresponding 'large-scale' peaks (in 2D) and (filament-type) saddles (in 3D), and build the corresponding fields $\boldsymbol{s}(\boldsymbol{r})$ (via `fft` using equation 24). This field allows us to assign a spin to each 'halo' at position $\boldsymbol{r}_\mathrm{h}$ and a closest saddle, $\boldsymbol{r}_\mathrm{s}$. Given the relative position $\boldsymbol{r}_\mathrm{h} - \boldsymbol{r}_\mathrm{s}$ as measured in the frame defined by the Hessian at $\boldsymbol{r}_\mathrm{s}$, we may project the direction of the spins, $\hat{\boldsymbol{s}} \equiv \boldsymbol{s}/s$ of all 'haloes' in the vicinity along the corresponding local cylindrical coordinate $(\boldsymbol{e}_R, \boldsymbol{e}_\phi, \boldsymbol{e}_z)$. We may then compute the one-point statistics of $\mu_z \equiv \hat{\boldsymbol{s}} \cdot \boldsymbol{e}_z$ *per octant*.

5.1.1 2D GRF fields spin flip

In two dimensions, the expectation is that the spin should be aligned or anti-aligned with e_z depending on each quadrant.

Let us first start with a set of 25 2D 2048^2 maps from a power spectrum with $n = -1/2$. The map is first smoothed with Gaussian filter of width $L_\mathrm{h} = 4$ pixels, and the positions of the peaks are identified. It is then smoothed again over $L_\mathrm{s} = 24$ pixels, exponentiated (in order to mimic the almost lognormal statistics of the evolved cosmic density field), and the corresponding Hessian and tidal fields are computed, together with the momentum map, which

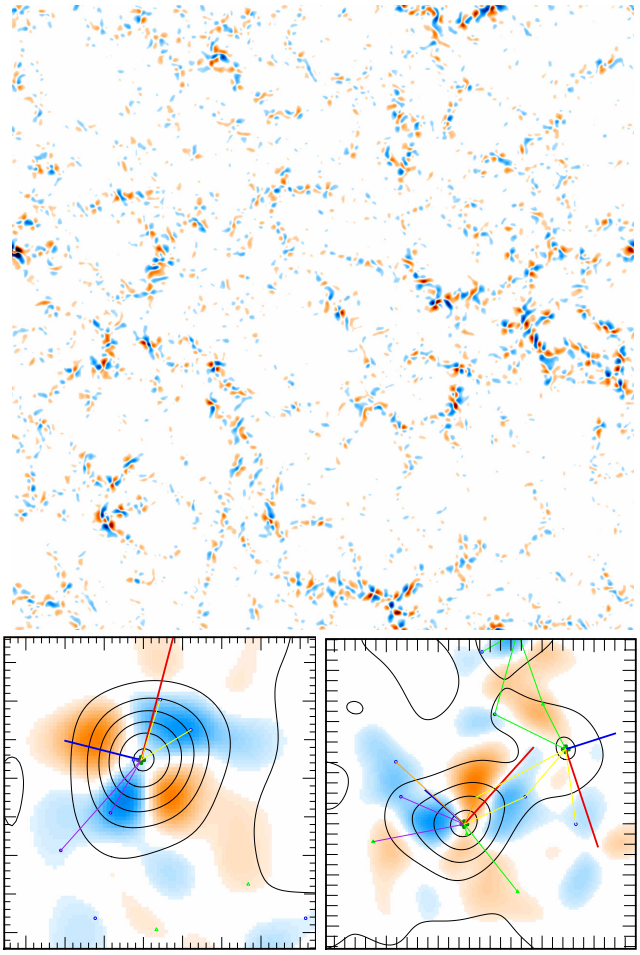


Figure 19. Top: example of spin map (colour coded by sign) generated following the prescription of Section 5.1.1. Bottom: the local frame (in red, long axis and blue) around a couple of 'saddle's. The black contours correspond to the density.

is thresholded above 1/30th of its highest value (see Fig. 19). The peaks of this second map are identified as 'saddles' for contrasts higher than 2.5. Fig. 20 shows that the average spin of 'haloes' in each quadrant is flipping from one quadrant to the next, with a statistically significant non-zero mean value in each quadrant.

5.1.2 3D GRF fields spin flip

Let us similarly consider a set of 20 three-dimensional 256^3 cubes from a power spectrum with $n = -2$. The cube is first smoothed with Gaussian filter of width $L_\mathrm{h} = 4$ pixels, and the positions of the peaks are identified. It is then smoothed again over $L_\mathrm{s} = 24$ pixels, exponentiated, and the corresponding Hessian and tidal fields are computed, together with the spin field, which is thresholded above 1/30th of its highest value. The saddle of this second cube are identified as for contrasts higher than 1. Only peaks closer than one smoothing length from the large-scale saddles are kept. The angle between their spin and the filament axis is computed and stored depending on the octant they belong to. In this section, the octants are numbered from 1 to 8 depending on the separation from the peak to the saddle $\boldsymbol{r} = (x, y, z)$: $x, y, z > 0$ (#1), $x < 0$ and $y, z > 0$ (#2), $x, y < 0$ and $z > 0$ (#3), $y < 0$ and $x, z > 0$ (#4), $z < 0$ and $x, y > 0$ (#5), $x, z < 0$ and $y > 0$ (#6), $x, y, z < 0$ (#7), and $y, z < 0$ and $x > 0$ (#8). Fig. 21 shows that, as expected, the component of the spin aligned with the filament axis is flipping sign from one octant to the other.

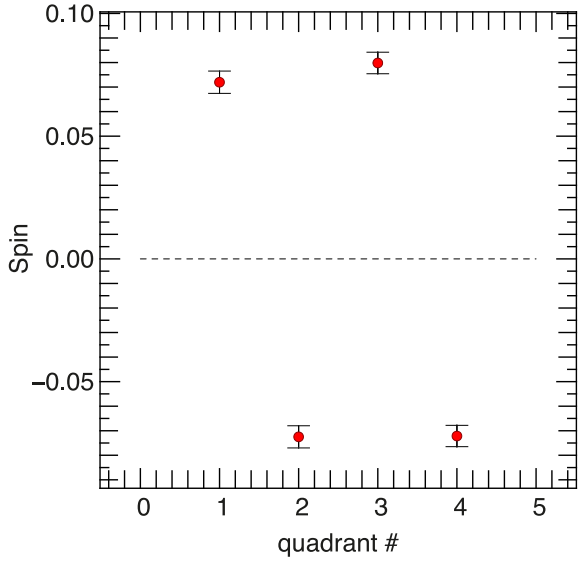


Figure 20. Alignment of 'spin' along $\boldsymbol{e}_z$ in two dimensions as a function of quadrant rank, clockwise. As expected, from one quadrant to the next, the spin is on average unambiguously flipping sign.

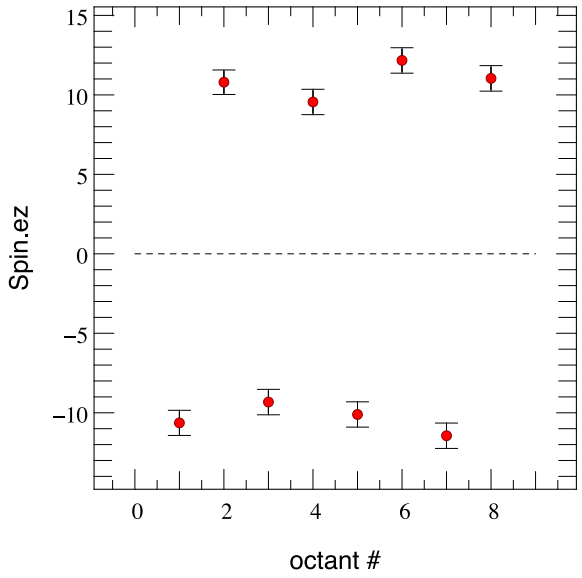


Figure 21. Alignment of the spin along the filamentary direction depending on the considered octant. As predicted by the theory, the z-component of the spin is flipping sign from one octant to the other.

5.2 Validation on dark matter simulations at $z = 0$

Let us now identify the Eulerian implication at redshift zero of the above sketched Lagrangian theory. For this, we must rely on N-body simulations. Hence, we now make use of the 43 million DM haloes detected at redshift zero in the Horizon-4π N-body simulation (Teyssier et al. 2009) to test some of the outcomes of the anisotropic tidal torque theory presented in this paper. This simulation contains 4096^3 DM particles distributed in a 2 h^{-1}Gpc periodic box and is characterized by the following ΛCDM cosmology:

$\Omega_m = 0.24$, $\Omega_\Lambda = 0.76$, $n = 0.958$, $H_0 = 73$ km s^{-1} Mpc^{-1} and $\sigma_8 = 0.77$ within one standard deviation of WMAP3 results (Spergel et al. 2003). The initial conditions were evolved non-linearly down to redshift zero using the adaptive mesh refinement code RAMSES (Teyssier 2002), on a 4096^3 grid. The motion of the particles was followed with a multigrid Particle-Mesh Poisson solver using a cloud-in-cell interpolation algorithm to assign these particles to the grid (the refinement strategy of 40 particles as a threshold for refinement allowed us to reach a constant physical resolution of 10 kpc, see the above-mentioned two references).

The Friend-of-Friend Algorithm (Huchra & Geller 1982) was used over 18^3 overlapping subsets of the simulation with a linking length of 0.2 times the mean inter-particular distance to define DM haloes. In the present work, we only consider haloes with more than 40 particles (the particle mass being 7.7×10^9 M$_\odot$). The mass dynamical range of this simulation spans about five decades.

The filament's direction is then defined via the global skeleton algorithm introduced by Sousbie, Colombi & Pichon (2009) and based on Morse theory. It defines the skeleton as the set of critical lines joining the maxima of the density field through saddle points following the gradient. In practice, Sousbie et al. (2009) define the peak and void patches of the density field as the set of points converging to a specific local maximum/minimum while following the field lines in the direction/opposite direction of the gradient. The skeleton is then the set of intersection of the void patches, i.e. the subset of critical lines connecting the saddle points and the local maxima of a density field and parallel to the gradient of the field. In practice, the ∼70 billion particles of the Horizon-4π were sampled on a 2048^3 Cartesian grid and the density field was smoothed using `mpsmooth` (Prunet et al. 2008) over a scale of $5\,h^{-1}$ Mpc corresponding to a mass of 1.9×10^{14} M$_\odot$. This cube was then divided into 6^3 overlapping subcubes and the skeleton was computed for each of these subcubes. It was then reconnected across the entire simulation volume to produce a catalogue of segments which locally defines the direction of the filaments.

Fig. 22 demonstrates that the spins of the 43 million dark haloes of the simulation obey the expected mass-dependent flip predicted by the theory presented in Section 4. On top of the alignment with the filament direction found, e.g. in Codis et al. (2012), haloes are shown to have a spin increasingly perpendicular to $\boldsymbol{e}_\phi$ at low-mass (red) and up to the critical mass ($\simeq 10^{12}$ M$_\odot$), while high-mass haloes have a spin parallel to the $\boldsymbol{e}_\phi$ direction. The transition from alignment to orthogonality occurs around $M_{\rm tr} \simeq 5 \times 10^{12}$ M$_\odot$.

Fig. 23 shows that the spins tend to be more aligned with the filament axis when getting closer to the saddle point. The alignment decreases from $\cos\theta = 0.511$ at $r \simeq 20$ Mpc h^{-1} to $\cos\theta = 0.506$ at $r < 1$ Mpc h^{-1}. This qualitative trend is in full agreement with the anisotropic tidal torque theory picture presented in Section 4 for which on average, spins are aligned with the filament axis in the plane of the saddle point and become misaligned when going away from this saddle point.

Fig. 24 displays the occupancy of haloes along the filaments. It appears that the higher the mass, the more concentrated they are far from the saddles. This is in good agreement with the halo mass gradient along the filaments described in Section 4.3.2.

Overall, the above GRF experiments as well as the re-analysis of the Horizon-4π N-body simulation seem consistent with the prediction of the theory presented in Sections 3 and 4. While the former demonstrates that interferences from neighbouring saddles do not wash out the tide correlations, the latter suggests that on the scales probed by this experiment, this Eulerian measure still captures features of the underlying Lagrangian theory.

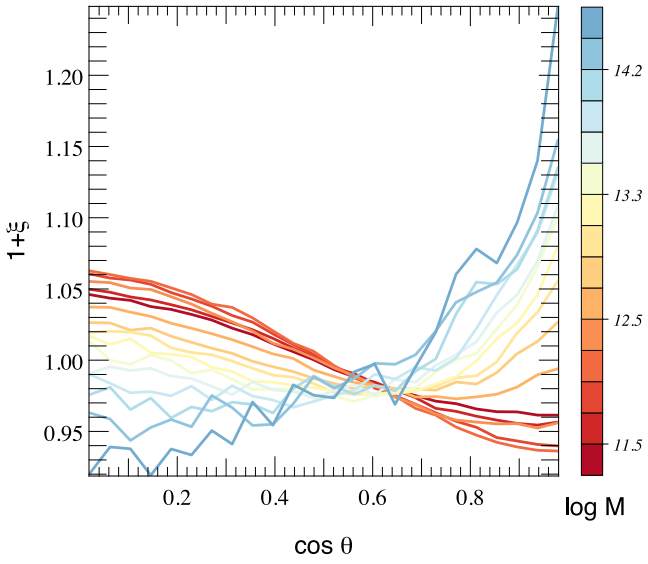


Figure 22. Alignment of 'spin' along $\boldsymbol{e}_\phi$ in the Horizon-4π simulation. The normalized histogram of the cosine of the angle between the spins and the closest filament's direction is displayed. Deviations from the $\xi = 0$ uniform distribution are detected and depends on the dark matter halo mass. Haloes have a spin aligned with the e_ϕ direction on average at low-mass (red) and perpendicular to it at larger mass (blue).

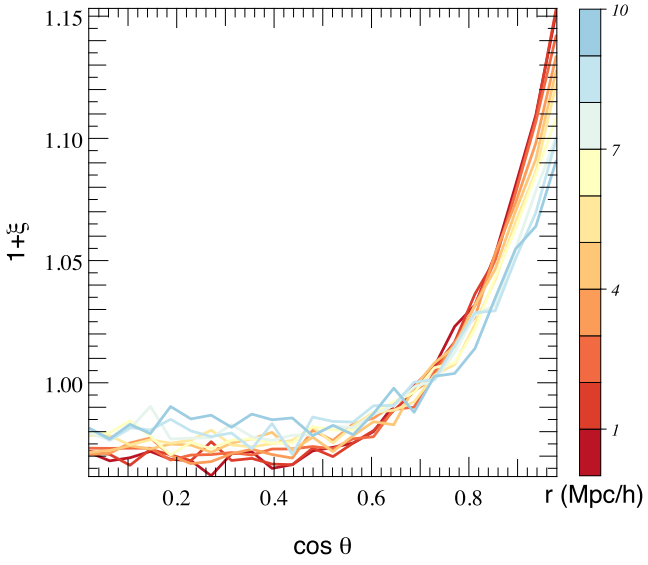


Figure 23. Alignment of the spins and the filaments in the Horizon-4π simulation as a function of the distance to the closest saddle point (from red – 0 – to blue – 10 Mpc h^{-1}). The alignment decreases with the distance to the saddle point as predicted by the anisotropic tidal torque theory model.

6 CONCLUSIONS AND PERSPECTIVES

TTT was revisited while focusing on an anisotropic peak-background-split in the vicinity of a saddle point. Such critical point captures as a point process the geometry of a typical filament embedded in a given wall (Pogosyan et al. 1998). The induced misalignment between the tidal tensor and the Hessian of the density simply explains the surrounding transverse and longitudinal point reflection-symmetric geometry of the spin distribution near filaments. This geometry of the spin field predicts in particular that less massive galaxies have their spin parallel to the filament, while more massive ones have their spin in the azimuthal direction. The

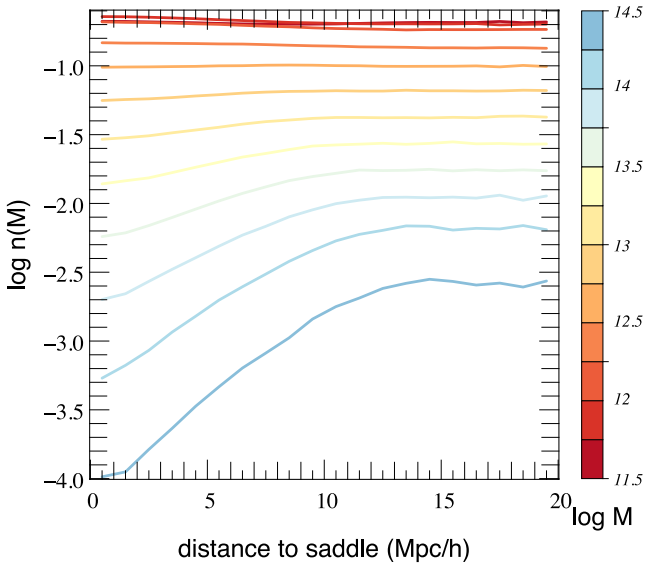


Figure 24. (log-)fraction of haloes of different mass (from red to blue in solar mass units) as a function of the distance to the saddle point in the Horizon-4π simulation. Low-mass haloes (red) lie almost uniformly along the filaments (with a small concentration – not clearly seen in logarithmic units – around the saddles due to consumption when going towards the nodes) while high-mass haloes (blue) are more concentrated far from the saddles.

corresponding transition mass follows from this geometry together with its scaling with the mass of non-linearity, in good agreement with measurements in simulations.

The main findings of this paper are: (i) galaxies form near filaments embedded in walls, and flow towards the nodes: this anisotropic environment produces the long wave modes on top of which galactic haloes pass the turnaround threshold; (ii) a typical filament is elongated and flattened: as a point process, it is therefore best characterized by its triaxial saddle points; (iii) the spin geometry is octupolar in the vicinity of the saddle point, displaying a point reflection symmetry; (iv) the mean spin field is parallel to the filament axis in the plane of the saddle point and becomes azimuthal away from it; (v) the constrained tidal torque theory presented in this paper allows to accurately predict the transition mass of the spin-filament alignment measured in simulations; (vi) this theory seems consistent with both GRF experiments and results from *N*-body simulations; (vii) a dual theory describes spin alignments in voids (see Appendix B).

6.1 Discussion

One of the striking features of this anisotropic extension of TTT is the induced quadrupolar point-symmetric flattened geometry of the spin distribution near a saddle point, which effectively scales down by 1 order of magnitude the transition mass away from the mass of non-linearity, in agreement with the measured scaling. The qualitative analysis derived from first principles in the vicinity of a given saddle point seems to hold when considering realizations of GRF, once proper account of the induced geometry near such points is taken care of. In effect, we have shown that the geometry of the saddle point provides a natural 'metric' (the local frame as defined by the Hessian at that saddle point) relative to which we can study the dynamical evolution of dark haloes along filaments. It should allow us to study how galactic feeding (via helicoidal cold flow; see Dubois et al. 2014) should vary with curvilinear coordinate along the filament. It was indeed found in that paper using hydrodynamical simulations that such flows were reaching galaxies in the so-called circum-galactic medium with velocities roughly parallel the polar axis. Taken at face value, such findings suggest that the flow feeding galaxies has significant helicity during that phase.

Another striking feature of this *Lagrangian* framework is that it captures naturally the arguably non-linear *Eulerian* process of spin flip via mergers. Recently, Laigle et al. (2015) showed that angular momentum generation of haloes could be captured in Eulerian space via the secondary advection of vorticity which the formation of the filament generates, whereas we show in this paper that it may also be described in Lagrangian space via the analysis of the anisotropic tides generated by the filament to be. No description is more fundamental than the other but are the two (Eulerian versus Lagrangian) sides of the same coin. The mapping between the two descriptions requires a reversible time integrator, such as the Zel'dovich approximation, which clearly limits its temporal validity to weakly non-linear scales. Our proxy for the spin, equation (3), is an approximation which seems to quantitatively capture the relevant physics. It is remarkable that such an (admittedly approximate) straightforward extension of TTT captures what seems to be the driving process of spin orientation acquisition and its initial evolution. It is also sticking that very simple closed form for the spin orientation distribution in the vicinity of the saddle point are available for this proxy.

Our theory here has focused on a two-scale process. Given the characteristics of ΛCDM hierarchical clustering, one can anticipate that this process occurs on several nested scales at various epochs – and arguably on various scales at the same epoch. The scenario we propose for the origin of this signal is, like the signal itself, relative to the linear scale involved in defining the filaments and as such, multiscale. It will hold as long as filaments are well defined in order to drive the local cosmic flow. In other words, one expects smaller scale filaments are themselves embedded in larger scale walls. The induced multiscale anisotropic flow transpires in the scaling of the transition mass with smoothing, as discussed in Codis et al. (2012).

Of course, we have here completely ignored the effect of feedback, which will play some – yet undefined – role in redistributing the cosmic pristine gas falling on to forming galaxies. Another issue would be to estimate for how long this entanglement between the large-scale dynamics and the kinematic properties of high redshift pervades, given the disruptions induced by feedback. What will be the effect of AGN feedback (Dubois et al. 2013; Prieto et al. 2015) on tidally biased secondary infall? Ocvirk, Pichon & Teyssier (2008) have also shown that at lower redshift, the so-called hot mode of accretion will kick in; how will hot flows wash out/disintegrate these ribbons? Given that they locally reflect the large-scale geometry, will the gas continue to flow-in along preferred directions (as does the DM; see e.g. Aubert, Pichon & Colombi 2004), or does the hot phase erase any anisotropy? Will the above-mentioned smaller scale non-linear dynamics eventually wash out any such trace?

6.2 Perspectives

One possibly significant shortcoming of the analysis is the proxy involved in using the Hessian of the density instead of the inertia tensor (though see Appendix A1). This is critical in order to retain a point process for the induced spin, but is achieved at the expense of having an adequate estimate for the *amplitude* of the spin, which is unfortunate because from the point of view of morphology, the dividing line between spirals and ellipticals is likely to be spin

amplitude. Let us none the less assume that, e.g. match to simulations or ansatz such as those described in Schäfer & Merkel (2012) will yield access to reasonable fit to spin amplitude and discuss briefly implications to galaxy formation within its cosmic web.

6.2.1 *Epoch of maximal spin advection?*

The inspection of hydrodynamical simulations (e.g. Codis et al. 2012, using tracer particles) shows that ribbon-like caustics feed the central galaxy along its spin axis from both poles. The gas flowing roughly parallel to the spin axis of the disc along both directions will typically impact the disc's circum-galactic medium and shock once more (as it did when it first reached the wall, and then the filaments, forming those above-mentioned ribbons), radiating away its vertical momentum (see Tillson et al. 2015). These ribbons are generated via the same winding/folding process as the protogalaxy, and represent the dominant source of secondary filamentary infall which feeds the newly formed galaxy with gas of well-aligned angular momentum.

Having computed the most likely spin (direction) as a function of position, it is therefore of interest to measure its covariant polar flux through a drifting forming galaxy.

From our knowledge of the spin distribution within the neighbourhood of a given saddle, we may then compute the rate of advected spin within some galactic volume $\mathcal{V} = S\Delta z$; it reads

$$\dot{\boldsymbol{s}} = \int \mathrm{d}^2\boldsymbol{S}\cdot\boldsymbol{v}\otimes\rho\boldsymbol{s} = \int_{\mathcal{V}} \mathrm{d}^3\boldsymbol{r}\ \nabla\cdot(\rho\boldsymbol{v}\otimes\boldsymbol{s})\ ,$$
$$\approx S\,[\rho\boldsymbol{v}\otimes\boldsymbol{s}]^{+}_{-} \approx S\Delta z\,\frac{\partial}{\partial z}\,(\rho v_z s_z)\ , \qquad (40)$$

where the last equality assumes that the advection is quasi-polar, and that the spin is mostly aligned with the filament. In equation (40), $\boldsymbol{v}$ is the gradient of the potential. Let us identify the curvilinear coordinate, $z_{\rm up}$, for which this flux is maximal:

$$z_{\rm up} = \mathrm{argmax}_z\,\frac{\partial}{\partial z}\,(\rho v_z s_z) = \left\{ z\,\middle|\,\frac{\partial^2}{\partial z^2}\,(\rho v_z s_z) = 0 \right\}. \qquad (41)$$

The coordinate $z_{\rm up}(\nu,\kappa_1,\kappa_2)$ characterizes the most active regions in the cosmic web for galactic spin-up. Focusing on the most likely saddle, the argument sketched in Section 4.3.2 allows us to assign a redshift-dependent spin-up mass, $\mathcal{M}_{\rm up}(Z)$, via equations (37) and (39). There could be an observational signature, e.g. in terms of the cosmic evolution of the SFR, as maximum spin-up corresponds to efficient pristine cold and dense gas accretion, which in turn induces consistent and steady star formation.

6.2.2 *Morphological type versus loci on web?*

The magnitude of the spin of galaxies could be taken as a proxy for morphological type. Indeed, Welker et al. (2014, 2015) have shown in cosmological hydrodynamical simulations that spin direction and galactic sizes where sensitive to the anisotropic environment. It is shown in particular that the magnitude of the spin of simulated galaxies increases steadily and aligns itself preferentially with the nearest filament when no significant merger occurs, in agreement with the first phase of the above described spin-up (see also Pichon et al. 2011). During that phase, the fraction of larger spirals should increase. In contrast, following Fig. 24, if we account for the fact that galactic morphology – the fraction of ellipticals, correlates with dark halo mass, it should then increase with distance to saddle.

In order to tackle such process theoretically, it would therefore be worthwhile to revisit Quinn & Binney (1992) in the context of this constrained theory of tidal torques and quantify how the dynamics of concentric shells are differentially biased by the tides of a saddle point. This would allow us to describe the whole timeline of anisotropic secondary infall.

6.2.3 *Implication for weak lensing?*

Weak lensing attempts to probe the statistics of the cosmic web between background galaxies – which shape is assumed to be uncorrelated – and the observer, while assuming that observed shape statistics reflects the deflection of light going through the intervening web. In view of Fig. 12, if we take as a proxy spin alignment for shape alignment, we can in principle compute the expectation of $\xi(\Delta\boldsymbol{r}) \equiv \langle \boldsymbol{s}(\boldsymbol{r})\cdot\boldsymbol{s}(\boldsymbol{r}')|\,\mathrm{skl}\,\rangle$ as a function of $\Delta\boldsymbol{r} = \boldsymbol{r} - \boldsymbol{r}'$. Calling $\delta\boldsymbol{s} = \boldsymbol{s} - \langle \boldsymbol{s}(\boldsymbol{r})|\,\mathrm{skl}\,\rangle$, we have $\xi(\Delta\boldsymbol{r}) = \langle \boldsymbol{s}(\boldsymbol{r})|\,\mathrm{skl}\rangle\cdot\langle \boldsymbol{s}(\boldsymbol{r}')|\,\mathrm{skl}\,\rangle + \langle\delta\boldsymbol{s}(\boldsymbol{r})\cdot\delta\boldsymbol{s}(\boldsymbol{r}')|\,\mathrm{skl}\,\rangle + 2\langle\delta\boldsymbol{s}(\boldsymbol{r})\cdot\boldsymbol{s}(\boldsymbol{r}')|\,\mathrm{skl}\,\rangle$. Let us just focus here on the first term, $\langle \boldsymbol{s}(\boldsymbol{r})|\,\mathrm{skl}\rangle\cdot\langle \boldsymbol{s}(\boldsymbol{r}')|\,\mathrm{skl}\,\rangle$. Given equation (25), we can compute it and find that it will typically be non-zero and vary significantly depending on both the magnitude and the orientation of $\Delta\boldsymbol{r}$. E.g. if $\Delta\boldsymbol{r}$ is off-axis along the filament, but if the pair is close to the saddle and $|\Delta\boldsymbol{r}|$ is small it will be positive (spins will align as they are both within coherent region of the saddle's tides), while if $|\Delta\boldsymbol{r}|$ is somewhat larger it will vanish (spins will be perpendicular). Conversely, if $\Delta\boldsymbol{r}$ is transverse to the filament and $|\Delta\boldsymbol{r}|$ is small, it will be positive, but if $|\Delta\boldsymbol{r}|$ is of the order of the size of one octant it will typically vanish again. The formalism presented in Section 3.1 can clearly be extended (while considering the joint three points statistics) to predict exactly all terms involved in $\xi(\Delta\boldsymbol{r})$ and quantify within this framework the effect of intrinsic alignments on the spin–spin two-point correlation. This is will be to topic of future work (Codis et al., in preparation).

ACKNOWLEDGEMENTS

This work is partially supported by the Spin(e) grants ANR-13-BS05-0005 (http://cosmicorigin.org) of the French *Agence Nationale de la Recherche* and by the ILP LABEX (under reference ANR-10-LABX-63 and ANR-11-IDEX-0004-02). CP thanks D. Lynden-bell for suggesting to tackle this problem and Churchill college for hospitality while this work was completed. Many thanks to J. Devriendt, A. Slyz, J. Binney, Y. Dubois, V. Desjacques, C. Laigle and S. Prunet for discussions about tidal torque theory, and to our collaborators of the Horizon project (http://projet-horizon.fr) for helping us produce the Horizon-4π simulation, F. Bouchet for allowing us to use the MAGIQUE3 supercomputer during commissioning, and to S. Rouberol for making it possible, and running the `horizon` cluster for us. SC and CP thank Lena for her hospitality when this work was initiated, and Eric for his help with some figures.

REFERENCES

Aragón-Calvo M. A., Yang L. F., 2014, MNRAS, 440, L46
Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007, ApJ, 655, L5
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bailin J., Steinmetz M., 2005, ApJ, 627, 647
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bernstein G. M., Norberg P., 2002, AJ, 124, 733
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002, MNRAS, 333, 501

Catelan P., Theuns T., 1996, MNRAS, 282, 436
Catelan P., Kamionkowski M., Blandford R. D., 2001, MNRAS, 320, L7
Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320
Codis S. et al., 2015, MNRAS, 448, 3391
Crittenden R. G., Natarajan P., Pen U.-L., Theuns T., 2001, ApJ, 559, 552
Croft R. A. C., Metzler C. A., 2000, ApJ, 545, 561
Danovich M., Dekel A., Hahn O., Teyssier R., 2012, MNRAS, 422, 1732
Doroshkevich A. G., 1970, Astrophysics, 6, 320
Dubois Y., Pichon C., Haehnelt M., Kimm T., Slyz A., Devriendt J., Pogosyan D., 2012, MNRAS, 423, 3616
Dubois Y., Pichon C., Devriendt J., Silk J., Haehnelt M., Kimm T., Slyz A., 2013, MNRAS, 428, 2885
Dubois Y. et al., 2014, MNRAS, 444, 1453
Forero-Romero J. E., Contreras S., Padilla N., 2014, MNRAS, 443, 1090
Gunn J. E., Gott J. R., III, 1972, ApJ, 176, 1
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489
Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649
Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, MNRAS, 347, 895
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Hirata C. M. et al., 2004, MNRAS, 353, 529
Hirata C. M., Mandelbaum R., Ishak M., Seljak U., Nichol R., Pimbblet K. A., Ross N. P., Wake D., 2007, MNRAS, 381, 1197
Hoyle F., 1949, Problems of Cosmical Aerodynamics. Central Air Documents, Office, Dayton, OH
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, A&A, 527, A26
Joachimi B., Semboloni E., Bett P. E., Hartlap J., Hilbert S., Hoekstra H., Schneider P., Schrabback T., 2013a, MNRAS, 431, 477
Joachimi B., Semboloni E., Hilbert S., Bett P. E., Hartlap J., Hoekstra H., Schneider P., 2013b, MNRAS, 436, 819
Klypin A., Shandarin S. F., 1993, ApJ, 413, 48
Laigle C. et al., 2015, MNRAS, 446, 2744
Lee J., Pen U., 2000, ApJ, 532, L5
Lee J., Pen U.-L., 2002, ApJ, 567, L111
Libeskind N. I., Hoffman Y., Forero-Romero J., Gottlöber S., Knebe A., Steinmetz M., Klypin A., 2013, MNRAS, 428, 2489
Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, MNRAS, 367, 611
Mandelbaum R. et al., 2011, MNRAS, 410, 844
Musso M., Sheth R. K., 2012, MNRAS, 423, L102
Ocvirk P., Pichon C., Teyssier R., 2008, MNRAS, 390, 1326
Paranjape A., Lam T. Y., Sheth R. K., 2012, MNRAS, 420, 1429
Paz D. J., Stasyszyn F., Padilla N. D., 2008, MNRAS, 389, 1127P
Peacock J. A., Heavens A. F., 1990, MNRAS, 243, 133
Peebles P. J. E., 1969, ApJ, 155, 393
Pichon C., Bernardeau F., 1999, A&A, 343, 663
Pichon C., Pogosyan D., Kimm T., Slyz A., Devriendt J., Dubois Y., 2011, MNRAS, pp 1739
Pogosyan D., Bond J. R., Kofman L., Wadsley J., 1998, in Colombi S., Mellier Y., Raban B., eds, Wide Field Surveys in Cosmology, Editions Frontieres, Paris, p. 61
Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F., Sousbie T., Colombi S., 2009, MNRAS, 396, 635
Porciani C., Dekel A., Hoffman Y., 2002, MNRAS, 332, 325
Prieto J., Jimenez R., Haiman Z., González R. E., 2015, MNRAS, 452, 784
Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R., Gottloeber S., 2008, ApJS, 178, 179
Quinn T., Binney J., 1992, MNRAS, 255, 729
Schaefer B. M., 2009, Int. J. Mod. Phys. D, 18, 173
Schäfer B. M., Merkel P. M., 2012, MNRAS, 421, 2751
Schneider M. D., Bridle S., 2010, MNRAS, 402, 2127
Schneider M. D., Frenk C. S., Cole S., 2012, J. Cosmol. Astropart. Phys., 5, 30
Sousbie T., Pichon C., Colombi S., Pogosyan D., 2008, MNRAS, 383, 1655
Sousbie T., Colombi S., Pichon C., 2009, MNRAS, 393, 457
Spergel D. N. et al., 2003, ApJS, 148, 175
Stewart K. R., Brooks A. M., Bullock J. S., Maller A. H., Diemand J., Wadsley J., Moustakas L. A., 2013, ApJ, 769, 74
Tempel E., Stoica R. S., Saar E., 2013, MNRAS, 428, 1827
Tenneti A., Singh S., Mandelbaum R., Matteo T. D., Feng Y., Khandai N., 2015, MNRAS, 448, 3522
Teyssier R., 2002, A&A, 385, 337
Teyssier R. et al., 2009, A&A, 497, 335
Tillson H., Devriendt J., Slyz A., Miller L., Pichon C., 2015, MNRAS, 449, 4363
Welker C., Devriendt J., Dubois Y., Pichon C., Peirani S., 2014, MNRAS, 445, L46
Welker C., Dubois Y., Devriendt J., Pichon C., 2015, preprint (arXiv:1510.4061)
White S. D. M., 1984, ApJ, 286, 38
White S. D. M., Tully R. B., Davis M., 1988, ApJ, 333, L45
Zel'dovich Y. B., 1970, A&A, 5, 84
Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747

APPENDIX A: A MULTISCALE THEORY

The proxy we take for the spin direction

$$s_i = \sum_{j,k,l} \epsilon_{ijk} H_{jl} T_{lk} \,, \tag{A1}$$

is, as mentioned in the main text, a (quadratic) approximation. First, because equation (1) is only valid in the linear regime (Porciani, Dekel & Hoffman 2002), but possibly more importantly because we take the Hessian as a proxy for the inertia tensor. In practice, recall that this approximation seems nevertheless to capture the essence of the processes at work in aligning spins with the large-scale structure given its ability to explain observed alignments through the comparison with simulations presented in Section 5.2 together with the measured mass transition. This suggests experimentally that it is indeed reasonable. Notwithstanding, while H_{ij} and I_{ij} locally share the same eigenframe, their amplitudes are different, leading to a different weighting of field configurations when computing ensemble averages such as in equation (14). It is therefore important to investigate this possible shortcoming further in this Appendix.

A1 More realistic spin proxies and peak

For this purpose, one can in principle (i) impose an additional peak constraint at the location where the spin is computed in order to impose that a protohalo will form there, and (ii) use more realistic spin proxies.

The peak constraint will typically be at a smaller scale than the filament's constraint, which requires building a two-scale theory and therefore increases significantly the complexity of the formalism. An additional difficulty with (ii) is that standard local proxies for the inertia tensor are highly non-linear and therefore require high-dimensional *numerical* integrations that are fairly difficult to implement in practice. For instance, considering the proxy that Schäfer & Merkel (2012) use to locally approximates the inertia tensor, we have

$$I_{ij} = \frac{M}{5}\begin{pmatrix} A_y^2+A_z^2 & 0 & 0 \\ 0 & A_z^2+A_x^2 & 0 \\ 0 & 0 & A_x^2+A_y^2 \end{pmatrix}, \tag{A2}$$

(in the frame of the Hessian) where the mass is $M = 4/3\pi A_x A_y A_z \rho_0 a_0^3$ and the semi-axes of the ellipsoid, A_i, are function of the eigenvalues of the Hessian (negative for a peak),

$$A_i = \sqrt{\frac{2\nu\sigma_2}{-\lambda_i}}. \tag{A3}$$

The traceless part of I_{ij} that is relevant for torques is then proportional to the traceless part of the inverse Hessian

$$\overline{I}_{ij} = \frac{2}{5}\nu\sigma_2 M \overline{H}^{-1}_{ij}. \tag{A4}$$

This introduces singular factors like $1/\sqrt{\det \boldsymbol{H}}$ in the expectation for $\langle s\rangle$. Such factors make the numerical evaluation of equation (24) more challenging as discussed in Section 2. We therefore postpone their evaluation in three dimensions to future work. Let us briefly investigate their implementation in two dimensions. In Appendix A2, we introduce a multiscale description while in Appendix A3 we take into account the proxy given by equation (A2) and we add an explicit peak condition.

A2 A 2D multiscale analysis

The evaluation of $r_\star$, see equation (20), requires to look for the scale that maximizes the spin amplitude. An improvement to the main text's approach is to investigate this issue in a two-scale theory where the scale of the central peak is larger than the scale of the halo. In this section, we propose to describe how to implement this multiscale theory and we show that this more accurate multiscale estimate for $r_\star$ yields, following the main text, a very similar value for the critical mass.

More specifically, let us smooth the fields in $\boldsymbol{r}_Y$ (the location of the spin) on R_h, an additional parameter, characterizing the halo's size and the fields in $\boldsymbol{r}_X$ on a different scale $R_\mathrm{s} > R_\mathrm{h}$ in order to impose a large-scale filament (a peak in 2D). If the one-point covariance matrix $\mathbf{C}_0$, does not change for a power-law power spectrum, the two-point covariance matrix $\mathbf{C}_\gamma$ does. In particular, all correlation functions ξ are now function of r, R_s and R_h. For instance for a power-law density power spectrum, $P(k) \propto k^n$,

$$\xi^{\Delta\Delta}_{\phi\phi}(r) = \left(\frac{R_\mathrm{s}^2+R_\mathrm{h}^2}{2R_\mathrm{s}R_\mathrm{h}}\right)^{-\frac{n+2}{2}}\left[\mathcal{F}_1^2 - \frac{1}{8}(n+2)\frac{r^2}{R_\mathrm{s}^2+R_\mathrm{h}^2}\mathcal{F}_2^3\right],$$

where $\mathcal{F}_i^j = {}_1F_1\left(n/2+i; j; -r^2/2(R_\mathrm{s}^2+R_\mathrm{h}^2)\right)$. The mean spin in $\boldsymbol{r}_Y$ is then given by the expectation of $\tilde{s}_z = \sum_{i,j,k}\epsilon_{ij3}\phi_{ik}x_{jk}$ given a peak on scale R_s in $\boldsymbol{r}_X$ as was computed in the main text. Compared to the main text, the only difference here is that we now also take into account the two-scale process through the two smoothing scales, R_s and R_h. The maximum spin magnitude as a function of the scale $R_\mathrm{s}/R_\mathrm{h}$ is then computed and displayed in Fig. A1. It appears that the spin magnitude is non-monotonic, peaking at $R_\mathrm{h} = 0.8R_\mathrm{s}$ which is very close to the value of $r_\star \approx 0.7R_\mathrm{s}$ (when top-hat smoothing are taken for both lengths).

Note that as we are computing here only the component of the spin along the filament, our magnitude plot does not contain the mass pre-factor, and our proxy for the moment of inertia, $\boldsymbol{I} \approx \boldsymbol{H}$, reflects only its orientation, but not its magnitude. We are able to argue, however,

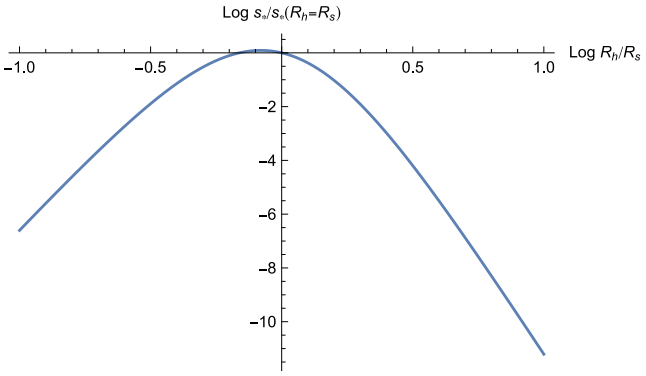


Figure A1. Maximum of spin as a function of the halo's scale R_h.

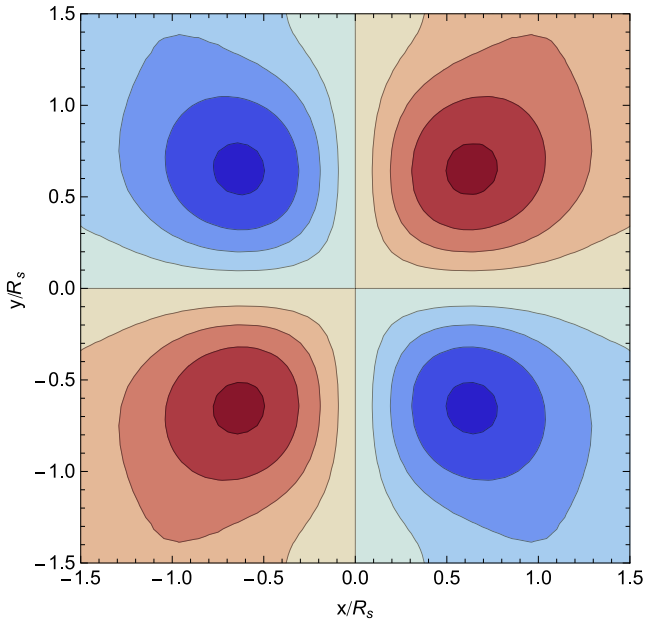


Figure A2. Mean spin computed via numerical integrations when (i) a central peak with geometry given by $\nu = 1$, $\lambda_1 = -1$, $\lambda_2 = -2$ is set, (ii) the inertia tensor is approximated by equation (A2), (iii) the scale of the central peak R_s is different from the scale of the halo R_h (here we take $R_h/R_s = 1/10$), (iv) there is a peak constraint at the location where the spin is computed with height $\nu = 5/2$, negative eigenvalues and zero gradient. The density power spectrum is a power law here with spectral index $n = 1/2$.

that Fig. A1 displays a maximum spin alignment with the filament's direction for some critical value of R_h – as it shows first an increase and then a fall in the z-component of the spin that is due to Hessian-tidal shear alignment.

A3 A 2D multiscale analysis with peak constraint

Adding a peak constraint at the location of the spin and taking into account the proxy given by equation (A2) is more difficult as it requires a numerical integration to account for the sign constraints on the eigenvalues. Notwithstanding this shortcoming, we will show now that in two dimensions, the adjunction of a peak constraint preserves both the qualitative picture (same geometry with four quadrants of opposite spin direction) as well as the typical scale for $r_\star$.

Fig. A2 indeed shows the numerical integration of the mean spin when (i) the inertia tensor is approximated by equation (A4) – where the mass is fixed by the smoothing length R_h –, (ii) the scale of the central peak R_s is different from the scale of the halo R_h (here we take $R_h/R_s = 1/10$), (iii) there is a peak constraint at the location where the spin is computed with height $\nu = 5/2$, negative eigenvalues and zero gradient. In short, the mean spin is now computed as

$$\langle \tilde{\mathbf{s}}|\mathrm{pk},\mathrm{pk}\rangle = \frac{\langle \tilde{\mathbf{s}}\,\det[\boldsymbol{H}]\,\Theta(-\lambda_i)\delta_{\mathrm{D}}(x-\nu)\delta_{\mathrm{D}}(x_i)|\mathrm{pk}\rangle}{\langle \det[\boldsymbol{H}]\,\Theta(-\lambda_i)\delta_{\mathrm{D}}(x-\nu)\delta_{\mathrm{D}}(x_i)|\mathrm{pk}\rangle}\,, \tag{A5}$$

where $\tilde{s}_i$ is defined as $\tilde{s}_i = \sum_{j,k,l}\epsilon_{ijk}H^{-1}_{jl}T_{lk}$ while the expectations $\langle\cdot|\mathrm{pk}\rangle$ are defined as conditional expectation to a central peak of geometry $\nu = 1$, $\lambda_1 = -1$, $\lambda_2 = -2$. The mean spin map is then obtained by numerical integration. Fig. A2 clearly shows that the four quadrants of opposite spin direction, as well as the size of these quadrants are preserved. This test strongly suggests that in two dimensions, improvements beyond the $\boldsymbol{I} \approx \boldsymbol{H}$ approximation do not change the global picture described in the main text.

Further developments, beyond the scope of this paper, could be to carry out the same analysis in three dimensions, also adding a peak constraint at the location where the spin is computed in order to impose the existence of a protohalo and use equation (A2) to define its inertia tensor. While the two-scale analysis is straightforward enough to implement, the adjunction of a peak constraint in three dimensions is much more tricky and requires in particular the computation of high-dimension numerical integrals (the results will not be analytic anymore) that are left for future investigations.

APPENDIX B: DUAL VOID THEORY

The theory presented in Section 4.1 is algebraic. Effectively no assumption has been made about the signs of the eigenvalues of the saddle we are considering. It is therefore also perfectly valid in the neighbourhood of a wall-type saddle in order to describe the spin alignments of dark haloes in that vicinity. At a qualitative level, Fig. 3 applies up to a sign: voids and wall saddles repel. It follows that the spins should rotate around the wall saddle to void axis and become parallel near the wall with a point symmetric change of polarity. This is indeed what equation (25) predicts and is shown on Fig. B1.

The statistical significance of these alignments is likely to be reduced as there are much fewer galaxies in voids and near wall saddles.

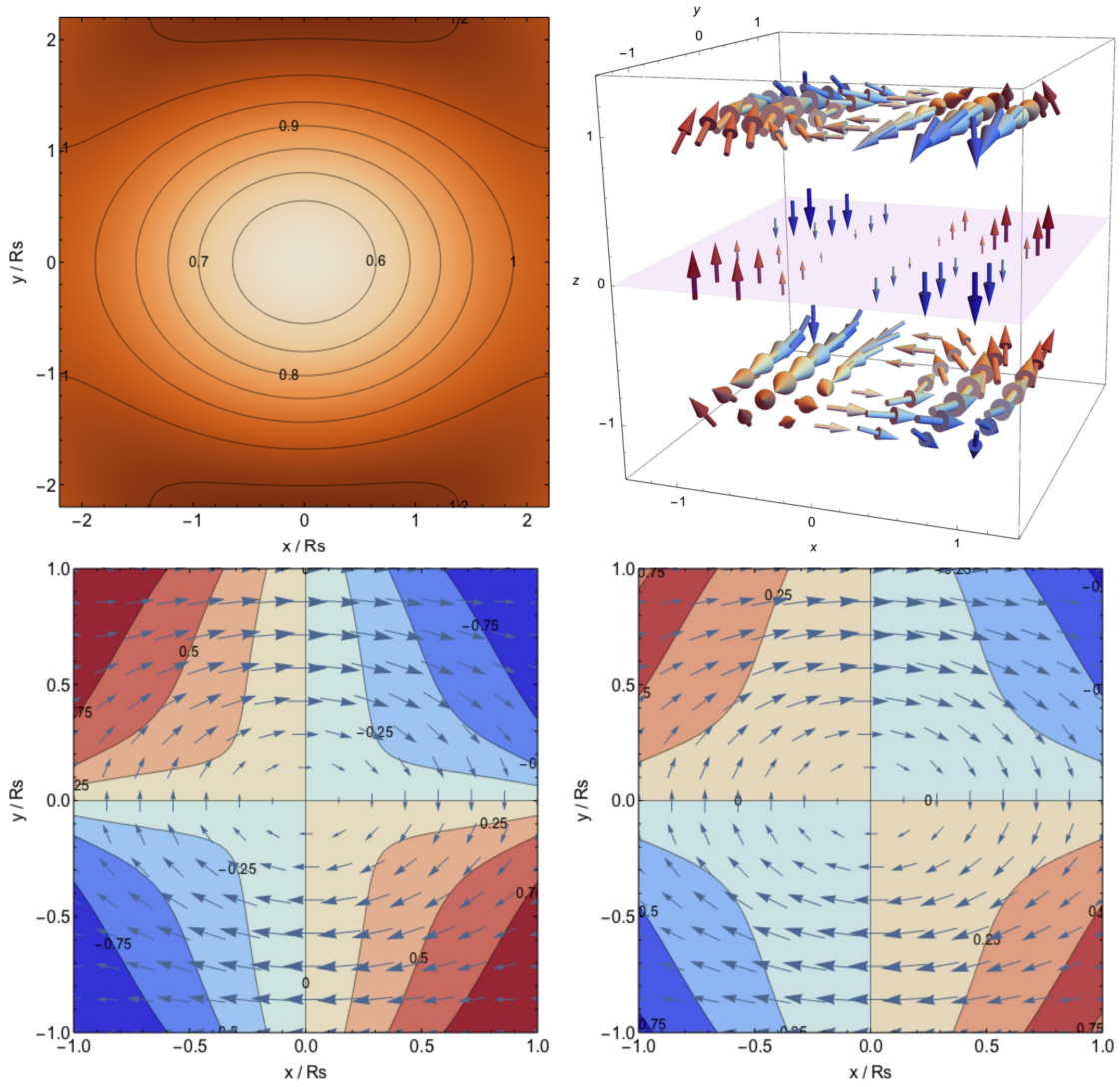


Figure B1. Top left-hand panel: mean density in the plane of the wall *Oxy* and centred on a wall-type filament with geometry $\nu = 0.5$, $\lambda_1 = 0.8$ along the x-axis, $\lambda_2 = 0.6$ along the y-axis and $\lambda_3 = -0.5$ along the z-axis for a power-law density power spectrum with spectral index $n = -2$. Contours are displayed from $\delta = 0.6$ to 1.2 as labelled. The filaments are clearly seen around $y = \pm 2R_s$. Top right-hand panel: mean spin colour coded by its projection along the normal to the wall. The spins are aligned with the normal in the plane of the wall and perpendicular to it when going outside the plane of the wall. Bottom panels: mean spin vectors at $z = 0.5R_s$ (left) and $z = 1R_s$ (right). Contours represent the orientation of the spin with regards to the normal to the wall from -1 (anti-aligned, red) to $+1$ (aligned, blue) through 0 (perpendicular).

APPENDIX C: TECHNICAL COMPLEMENTS

C1 Codes for density and spin in 2/3D

The expression for the 2 and 3D spin statistics (mean and variance) for scale-invariant power spectra are available both as a mathematica package (http://www.iap.fr/users/pichon/spin/code/ATTT.m), and a mathematica notebook (http://www.iap.fr/users/pichon/spin/code/ATTT-package.nb). The following functions are provided: `δ2D, spin2D, var2D, δ3D, spin3D,` which correspond, resp. to the 2D density, spin, its variance, and in 3D the density and the spin for scale-invariant power spectra of index n as a function of position $r, \theta, (\phi)$ and the geometry of the peak (resp. saddle) $\nu, \lambda_1, \lambda_2, (\lambda_3)$. Compiled versions are also provided.

C2 Correlation functions for power-law spectra

The 2D correlation functions defined in equations (5) and (9) can be analytically obtained for density power-law power spectrum $P_k(k) \propto k^n$ for which the scale parameter is $\gamma = \sqrt{(n+2)/(n+4)}$:

$$\xi^{\Delta\Delta}_{\phi\phi}(r) = \mathcal{F}_1^2 - \frac{1}{16}(n+2)r^2\mathcal{F}_2^3,$$

$$
\begin{aligned}
\xi^{\Delta\Delta}_{\phi x}(r) &= \gamma\left(\frac{(n+4)}{16}r^2\mathcal{F}^3_3 - \mathcal{F}^2_2\right),\\
\xi^{\Delta\Delta}_{xx}(r) &= \mathcal{F}^2_3 - \frac{1}{16}(n+6)r^2\mathcal{F}^3_4,\\
\xi^{\Delta+}_{\phi\phi}(r) &= -\frac{1}{32}(n+2)r^2\mathcal{F}^3_2,\\
\xi^{\Delta+}_{\phi x}(r) &= \frac{\gamma}{32}(n+4)r^2\mathcal{F}^3_3,\\
\xi^{\Delta+}_{xx}(r) &= -\frac{1}{32}(n+6)r^2\mathcal{F}^3_4,\\
\xi^{\times\times}_{\phi\phi}(r) &= \frac{1}{8}\left(4\mathcal{F}^2_1 - 3\mathcal{F}^3_1\right),\\
\xi^{\times\times}_{\phi x}(r) &= -\frac{\gamma}{8}\left(4\mathcal{F}^2_2 - 3\mathcal{F}^3_2\right)\\
\xi^{\times\times}_{xx}(r) &= \frac{1}{8}\left(4\mathcal{F}^2_3 - 3\mathcal{F}^3_3\right),
\end{aligned}
\tag{C1}
$$

where r is in units of the smoothing length and $\mathcal{F}^j_i = {}_1F_1\left(n/2+i; j; -r^2/4\right)$, with ${}_1F_1$ the Hypergeometric functions of the first kind. Some of those correlation functions are plotted in Fig. 5.

The 3D correlation functions can similarly be obtained for power-law power spectrum $P_k(k) \propto k^n$(some of those correlations are plotted in Fig. 10) for which the scale parameter is $\gamma = \sqrt{(n+3)/(n+5)}$ and we define $\mathcal{G}^j_i = {}_1F_1\left(n+i/2; j/2; -r^2/4\right)$:

$$
\begin{aligned}
\xi^{\Delta\Delta}_{\phi\phi}(r) &= \frac{-32(n-1)\mathcal{G}^3_1 - \left((2-4n)r^2 + r^4 - 32\right)\mathcal{G}^1_{-1} + (n-2)\left(2nr^2 - r^4 + 32\right)\mathcal{G}^3_{-1}}{2\left(n^2-1\right)r^2},\\
\xi^{\Delta\Delta}_{\phi x}(r) &= \frac{\Gamma\left(\frac{n+1}{2}\right)\left(32(n+1)\mathcal{G}^3_3 + \left(-2(2n+3)r^2 + r^4 - 32\right)\mathcal{G}^1_1 + n\left(-2(n+2)r^2 + r^4 - 32\right)\mathcal{G}^3_1\right)}{8r^2\sqrt{\Gamma\left(\frac{n+3}{2}\right)}\sqrt{\Gamma\left(\frac{n+7}{2}\right)}},\\
\xi^{\Delta\Delta}_{xx}(r) &= \frac{-32(n+3)\mathcal{G}^3_5 + \left(2(2n+7)r^2 - r^4 + 32\right)\mathcal{G}^1_3 + (n+2)\left(2(n+4)r^2 - r^4 + 32\right)\mathcal{G}^3_3}{2(n+3)(n+5)r^2},\\
\xi^{\Delta+}_{\phi\phi}(r) &= \frac{\left(4n\left(r^2+3\right) - \left(r^2+8\right)r^2 - 40\right)\mathcal{G}^1_{-1} + 16(n-1)\mathcal{G}^3_1 + (n-2)\left(2(n-3)r^2 - r^4 - 28\right)\mathcal{G}^3_{-1}}{2\left(n^2-1\right)r^2},\\
\xi^{\Delta+}_{\phi x}(r) &= \frac{\Gamma\left(\frac{n+1}{2}\right)\left(-16(n+1)\mathcal{G}^3_3 + \left(-4n\left(r^2+3\right) + r^4 + 16\right)\mathcal{G}^1_1 + n\left(-2(n-1)r^2 + r^4 + 28\right)\mathcal{G}^3_1\right)}{8r^2\sqrt{\Gamma\left(\frac{n+3}{2}\right)}\sqrt{\Gamma\left(\frac{n+7}{2}\right)}},\\
\xi^{\Delta+}_{xx}(r) &= \frac{16(n+3)\mathcal{G}^3_5 + \left(4(n+2)r^2 + 12n - r^4 + 8\right)\mathcal{G}^1_3 + (n+2)\left(2(n+1)r^2 - r^4 - 28\right)\mathcal{G}^3_3}{2(n+3)(n+5)r^2},\\
\xi^{\times\times}_{\phi\phi}(r) &= \frac{\left((n-2)r^2\left(r^2+10\right) - 48\right)\mathcal{G}^3_{-1} + \left(-2(n-6)r^2 + r^4 + 48\right)\mathcal{G}^1_{-1}}{\left(n^2-1\right)r^4},\\
\xi^{\times\times}_{\phi x}(r) &= -\frac{\Gamma\left(\frac{n+1}{2}\right)\left(\left(nr^2\left(r^2+10\right) - 48\right)\mathcal{G}^3_1 + \left(-2(n-4)r^2 + r^4 + 48\right)\mathcal{G}^1_1\right)}{4r^4\sqrt{\Gamma\left(\frac{n+3}{2}\right)}\sqrt{\Gamma\left(\frac{n+7}{2}\right)}},\\
\xi^{\times\times}_{xx}(r) &= \frac{\left((n+2)r^2\left(r^2+10\right) - 48\right)\mathcal{G}^3_3 + \left(-2(n-2)r^2 + r^4 + 48\right)\mathcal{G}^1_3}{(n+3)(n+5)r^4}.
\end{aligned}
$$

C3 Correlation functions for LCDM spectra

The same ξ correlation functions can also be computed for a ΛCDM power spectrum using Bardeen et al. (1986) and equation (9). The corresponding functions are shown on Fig. C1 for a Gaussian smoothing length of $R_s = 5\mathrm{Mpc}\,h^{-1}$ and a WMAP-7 cosmology. Note that those correlation functions are quite similar to $n = -2$ power-law power spectrum (see Fig. 10). Given these correlations, it would be straightforward to compute the corresponding spin.

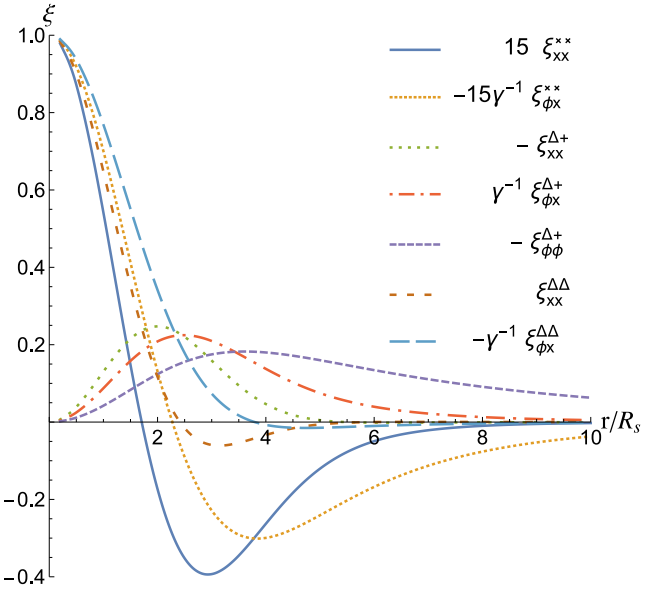


Figure C1. Two-point correlation functions as a function of the separation r in units of the smoothing length $R_s = 5$ Mpc h^{-1} for a ΛCDM power spectrum.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **448,** 3391–3404 (2015) doi:10.1093/mnras/stv231

Intrinsic alignment of simulated galaxies in the cosmic web: implications for weak lensing surveys

S. Codis,[1,2]★ R. Gavazzi,[1,2] Y. Dubois,[1,2] C. Pichon,[1,2] K. Benabed,[1,2] V. Desjacques,[3] D. Pogosyan,[4] J. Devriendt[5] and A. Slyz[5]

[1]*CNRS, UMR7095, Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, F-75014 Paris, France*
[2]*Sorbonne Universités, UPMC Univ. Paris 06, UMR7095, Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, F-75014 Paris, France*
[3]*Département de Physique Théorique, Université de Genève, 24 quai Ernest Ansermet, CH-1211 Genève, Switzerland*
[4]*Department of Physics, University of Alberta, 11322-89 Avenue, Edmonton, Alberta T6G 2G7, Canada*
[5]*Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*

Accepted 2015 February 3. Received 2015 January 30; in original form 2014 June 18

ABSTRACT

The intrinsic alignment of galaxy shapes (by means of their angular momentum) and their cross-correlation with the surrounding dark matter tidal field are investigated using the 160 000, $z = 1.2$ synthetic galaxies extracted from the high-resolution cosmological hydrodynamical simulation HORIZON-AGN. One- and two-point statistics of the spin of the stellar component are measured as a function of mass and colour. For the low-mass galaxies, this spin is *locally* aligned with the tidal field 'filamentary' direction while, for the high-mass galaxies, it is *perpendicular* to both filaments and walls. The bluest galaxies of our synthetic catalogue are more strongly correlated with the surrounding tidal field than the reddest galaxies, and this correlation extends up to $\sim 10\,h^{-1}$ Mpc comoving distance. We also report a correlation of the projected ellipticities of blue, intermediate-mass galaxies on a similar scale at a level of 10^{-4} which could be a concern for cosmic shear measurements. We do not report any measurable intrinsic alignments of the reddest galaxies of our sample. This work is a first step towards the use of very realistic catalogue of synthetic galaxies to evaluate the contamination of weak lensing measurement by the intrinsic galactic alignments.

Key words: gravitational lensing: weak – methods: numerical – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

For the last two decades, weak gravitational lensing has emerged as one of the most promising cosmological probes of the dark matter and dark energy contents of the Universe, culminating in the design of several large surveys like Dark Energy Survey (DES),[1] *Euclid* (Laureijs et al. 2011) or Large Synoptic Survey Telescope (LSST).[2]

As the statistical power of weak lensing surveys is ramping up, more and more attention has to be paid for the control of systematic effects. Among the critical astrophysical sources of errors is the problem of the intrinsic alignments (IA) of galaxies. The fundamental assumption upon which galaxies are randomly aligned in the absence of a shear signal that is coherent on the scales of several arcminutes is likely to break down for pairs of galaxies observed at close angular distances (through direct gravitational interactions or as a result of the same local tidal field they live in). Much effort has thus been made to control the level of IA of galaxies as a potential source of systematic errors in weak gravitational lensing measurements (e.g. Croft & Metzler 2000; Heavens, Refregier & Heymans 2000; Hirata & Seljak 2004), although some techniques have been proposed to mitigate their nuisance by making extensive use of photometric redshifts (e.g. Bridle & King 2007; Joachimi & Schneider 2008, 2010; Joachimi & Bridle 2010; Kirk, Bridle & Schneider 2010; Blazek et al. 2012).

Direct measurements of the alignment of the projected light distribution of galaxies in wide field imaging data seem to agree on a contamination at a level of a few per cent in the shear correlation functions, although the amplitude of the effect depends on the depth of observations (stronger for shallower surveys), the amount of redshift information and the population of galaxies considered (in the sense that red galaxies seem to show a strong intrinsic projected shape alignment signal whereas observations only place upper limits in the amplitude of the signal for blue galaxies; Bernstein & Norberg 2002; Brown et al. 2002; Lee & Pen 2002; Heymans et al. 2004; Hirata et al. 2004, 2007; Mandelbaum et al. 2006, 2011; Joachimi et al. 2011, 2013a). Direct observations of the alignment between the spin and the tidal tensor eigenvectors have also been

★ E-mail: codis@iap.fr
[1] http://www.darkenergysurvey.org
[2] http://www.lsst.org

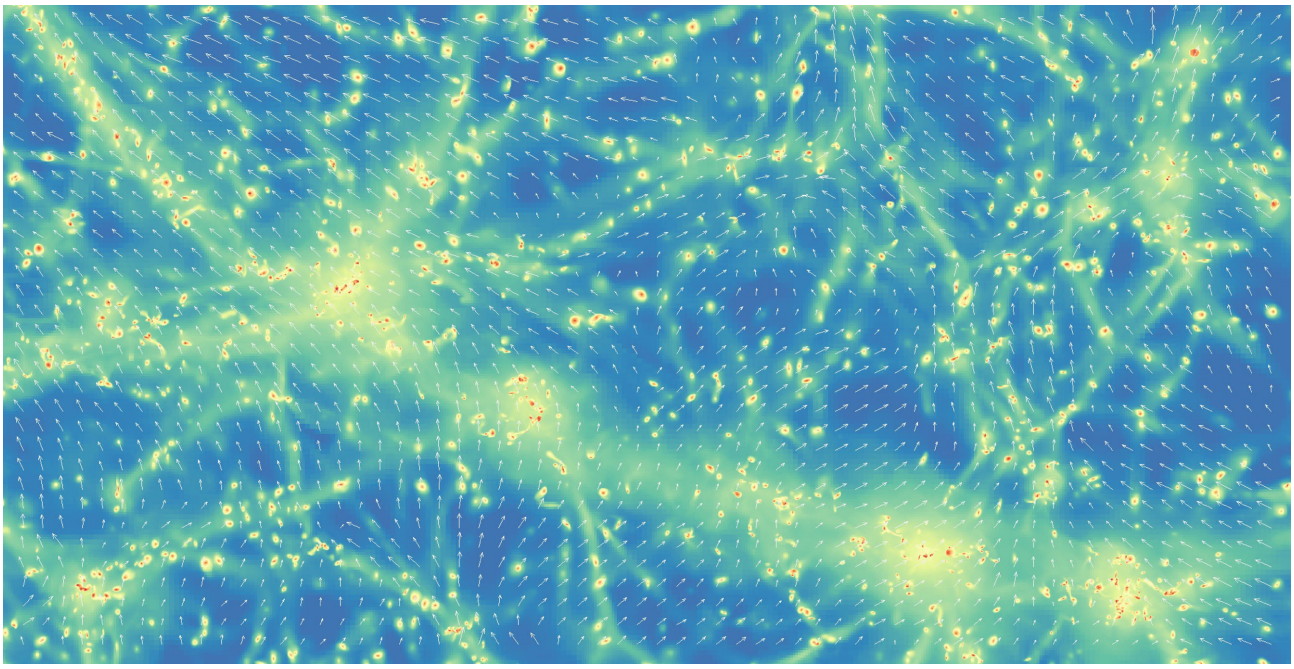


Figure 1. The $\boldsymbol{e}_1$ eigenvector (white arrows) of the tidal field within a slice of 25 h^{-1} Mpc comoving in depth and 12.5 h^{-1} Mpc comoving horizontally together with the gas density (from blue to red) within the HORIZON-AGN simulation at $z = 1.2$. As expected, $\boldsymbol{e}_1$ statistically follows the filaments.

carried out: the first attempt by Lee & Pen (2002) studied the correlations between the disc orientation of the galaxies from the Tully catalogue and the shear reconstructed from the Point Source Catalogue Redshift Survey and confidently rejected the hypothesis of randomness. More recently, Lee et al. (2013) detected some correlations between the spin and the intermediate eigenvector of the tidal tensor and found that galactic spins were also preferentially perpendicular to the major principal axis but this signal remains weak.

Cosmological numerical simulations are a natural way of further refining our models of IA. The imprint of the large-scale dynamics on to the shapes and spins of galaxies has been extensively studied using dark matter (DM) simulations (Aubert, Pichon & Colombi 2004; Bailin & Steinmetz 2005; Aragón-Calvo et al. 2007; Hahn et al. 2007b; Lee et al. 2008; Paz, Stasyszyn & Padilla 2008; Sousbie et al. 2008; Codis et al. 2012; Laigle et al. 2015; Forero-Romero, Contreras & Padilla 2014, among others). One can also mention a numerical study of the alignment between halo spin and tidal tensor by Porciani, Dekel & Hoffman (2002), who predicted its orthogonality with the major principal axis but also found that galactic spins must have lost their initial alignment with the tidal tensor predicted by tidal-torque theory (TTT), and by Codis et al. (2012). However, given the complex dependency on the physical properties of the galaxies seen in the observation, it is probably difficult to rely on DM-only numerical simulations as the sole resort to predict and control IA for weak lensing applications despite some success with the addition of halo model or semi-analytical models (e.g. Schneider & Bridle 2010; Joachimi et al. 2013b).

The advent of hydrodynamical cosmological simulations is arguably the best way forward to make better predictions on the complex relation between halo shape and spin and galaxy shape and spin. Local studies of the relation between DM and baryonic spins or inertia tensors have been conducted to measure the degree of alignment between the inertia tensor of the DM halo and the stellar component (not weighted by luminosity; Hahn, Teyssier & Carollo 2010; Tenneti et al. 2014). They find a typical excursion of the misalignment angle of 30°–10° for haloes ranging in mass from 10^{10} to 10^{14} h^{-1} $M_\odot$.

More recently, Dubois et al. (2014) measured the alignment of galaxy spins with the large-scale filamentary network in the HORIZON-AGN simulation, a state-of-the-art hydrodynamical simulation which produced synthetic galaxies displaying morphological diversity by redshift $z = 1.2$. Dubois et al. found that simulated galaxies have a spin which is either parallel to their neighbouring filament for low-mass, disc-dominated blue galaxies, or perpendicular to it for high-mass, velocity-dispersion-dominated red galaxies, the rapid reorientation of the latter massive galaxies being due to mergers. This suggests a scenario in which galaxies form in the vorticity-rich neighbourhood of filaments, and then migrate towards the nodes of the cosmic web, converting their orbital momentum into spin. The inherently anisotropic nature of the large-scale structure (filaments and walls) and its complex imprint on the shape and spin of galaxies (see also Pichon et al. 2014) may prevent isotropic approaches from making accurate predictions and suggest that the imprint of IA might be more severe for higher order statistics of the cosmic shear signal and definitely not addressable with simple prescriptions for the relation between the spin or inertia tensor of haloes and galaxies.

We thus propose here to extend the work of Dubois et al. (2014) by bringing the findings of the HORIZON-AGN simulation (see Fig. 1) closer to the framework of weak lensing observables. In particular we will measure the correlations between galactic spins and their surrounding tidal field (related to the so-called GI term of Hirata & Seljak 2004) and the correlations between spins themselves (related to the so-called II term of the same reference). We will also exhibit the variation of this quantity with the mass and colour of our galaxies, looking for populations where the IA effect is particularly severe or reduced. Our main finding is an excess alignment between the bluest galaxies of our synthetic catalogue, and no detectable alignment for the reddest ones. This conclusion

is in apparent contradiction with the works cited above, and we will discuss how this can be explained mainly by selection effects (on mass range, redshifts of catalogue, etc.).

Throughout this work, we use the stellar spin as a proxy for the ellipticity of galaxies, without attempting to project galactic ellipticities perpendicular to a given sightline. This choice differs from other authors who rather considered the inertia tensor of the stellar mass (see e.g. Tenneti et al. 2014). We believe spin can give a complementary insight on the apparent luminosity-weighted projected morphology of a generally star-forming galaxy. In addition, as one gets closer to the resolution limit of the simulation, we believe that the reliability of the simulated spin will hold longer than the reliability of the overall shape of the stellar component. Although we mainly focus on 3D quantities that are better suited to quantify the physical degree of IA in our simulation, we give some guidelines for inferring projected quantities, in the usual formalism of weak lensing.

The paper is organized as follows. Section 2 presents the HORIZON-AGN simulation and describes our method for measuring spin and inertia tensor. It also illustrates how much the shape inferred from spin can be favourably compared to the shape inferred from the inertia tensor of stars, hence supporting our choice of using the spin as a proxy for the ellipticity of galaxies. Section 3 defines how intrinsic alignments are quantified and where they contaminate the weak lensing observables. In Section 4, we present the cross-correlation between the principal axes of the tidal tensor and the spin vector as a function of distance and further show the zero-lag one-point probability distribution function (PDF) of this angle. Section 5 investigates the spin–spin two-point correlation as a function of separation and the projected ellipticity two-point correlation function. Section 6 checks that grid locking effects do not dominate the measurements. We finally conclude in Section 7 and discuss briefly how the statistics depends on the synthetic colours of galaxies. We also sketch how our findings can be cast into predictions on the contamination of weak lensing by IA. Appendix A studies the corresponding alignments for DM haloes.

2 THE SYNTHETIC UNIVERSE

Let us shortly describe the HORIZON-AGN simulation (Section 2.1; see Dubois et al. 2014 for more details) and explain how galaxy properties are extracted out of it (Section 2.2).

2.1 The HORIZON-AGN simulation

A standard Λ cold dark matter (ΛCDM) cosmology compatible with the 7-year *Wilkinson Microwave Anisotropy Probe* (*WMAP*-7) cosmology (Komatsu et al. 2011) is adopted, with total matter density $\Omega_\mathrm{m} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, amplitude of the matter power spectrum $\sigma_8 = 0.81$, baryon density $\Omega_\mathrm{b} = 0.045$, Hubble constant $H_0 = 70.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $n_\mathrm{s} = 0.967$. The HORIZON-AGN simulation has been run with 1024^3 DM particles in a $L_\mathrm{box} = 100\,h^{-1}$ Mpc box, so as to obtain a DM mass resolution of $M_\mathrm{DM,res} = 8 \times 10^7\,\mathrm{M}_\odot$. The adaptive mesh refinement (AMR) code RAMSES (Teyssier 2002) has been used to run the simulation with an initial mesh refinement of up to $\Delta x = 1$ kpc (seven levels of refinement). The refinement scheme follows a quasi-Lagrangian criterion: if the number of DM particles in a cell is more than eight, or if the total baryonic mass in a cell is eight times the initial DM mass resolution, a new refinement level is triggered.

A Sutherland & Dopita (1993) model is used to allow gas cooling by means of H and He cooling down to 10^4 K with a contribution from metals. Following Haardt & Madau (1996), heating from a uniform ultraviolet (UV) background takes place after redshift $z_\mathrm{reion} = 10$. We model metallicity as a passive variable for the gas that varies according to the injection of gas ejecta during supernovae explosions and stellar winds. A Schmidt law is used to model star formation: $\dot{\rho}_* = \epsilon_* \rho / t_\mathrm{ff}$, where $\dot{\rho}_*$ is the star formation rate density, $\epsilon_* = 0.02$ (Kennicutt 1998; Krumholz & Tan 2007) the constant star formation efficiency and t_ff the local free-fall time of the gas. We allow star formation where the gas hydrogen number density exceeds $n_0 = 0.1\,\mathrm{H\,cm^{-3}}$ according to a Poisson random process (Rasera & Teyssier 2006; Dubois & Teyssier 2008) with a stellar mass resolution of $M_\mathrm{*,res} = \rho_0 \Delta x^3 \simeq 2 \times 10^6\,\mathrm{M}_\odot$.

We model stellar feedback using a Salpeter (1955) initial mass function with a low-mass (high-mass) cut-off of $0.1\,\mathrm{M}_\odot$ ($100\,\mathrm{M}_\odot$). In particular, the mechanical energy from Type II supernovae and stellar winds follows the prescription of STARBURST99 (Leitherer et al. 1999, 2010), and the frequency of Type Ia supernovae explosions is taken from Greggio & Renzini (1983).

Active galactic nuclei (AGN) feedback is modelled according to Dubois et al. (2012). A Bondi–Hoyle–Lyttleton accretion rate on to black holes is used $\dot{M}_\mathrm{BH} = 4\pi\alpha G^2 M_\mathrm{BH}^2 \bar{\rho} / (\bar{c}_\mathrm{s}^2 + \bar{u}^2)^{3/2}$, where M_BH is the black hole (BH) mass, $\bar{\rho}$ is the average gas density, $\bar{c}_\mathrm{s}$ is the average sound speed, $\bar{u}$ is the average gas velocity relative to the BH velocity and α is a dimensionless boost factor with $\alpha = (\rho/\rho_0)^2$ when $\rho > \rho_0$ and $\alpha = 1$ otherwise (Booth & Schaye 2009) in order to account for our inability to capture the colder and higher density regions of the interstellar medium. The effective accretion rate on to BHs is capped at the Eddington accretion rate: $\dot{M}_\mathrm{Edd} = 4\pi G M_\mathrm{BH} m_\mathrm{p} / (\epsilon_\mathrm{r} \sigma_\mathrm{T} c)$, where σ_T is the Thompson cross-section, c is the speed of light, m_p is the proton mass and ϵ_r is the radiative efficiency, assumed to be equal to $\epsilon_\mathrm{r} = 0.1$ for the Shakura & Sunyaev (1973) accretion on to a Schwarzschild BH. Two different modes of AGN feedback are accounted for, the *radio* mode operating when $\chi = \dot{M}_\mathrm{BH} / \dot{M}_\mathrm{Edd} < 0.01$ and the *quasar* mode active otherwise. More details are given in Dubois et al. (2014).

2.2 Data analysis

2.2.1 *Galaxy catalogue*

Galaxies are identified with the ADAPTAHOP finder (Aubert et al. 2004), which relies directly on the distribution of star particles to construct the catalogue of galaxies. 20 neighbours are used to compute the local density of each particle. A local threshold of $\rho_\mathrm{t} = 178$ times the average total matter density is applied to select relevant densities. Note that the galaxy population does not depend sensitively on the exact value chosen for this threshold. Our specific choice reflects the fact that the average density of galaxies located at the centre of galaxy clusters is comparable to that of their host. The force softening (minimum size below which substructures are treated as irrelevant) is of $\sim$10 kpc. Only galactic structures identified with more than 50 star particles are included in the mock catalogues. This enables a clear identification of galaxies, including those in the process of merging. A galaxy catalogues with $\sim$165 000 objects is produced at $z = 1.2$ with masses between 1.7×10^8 and $1.4 \times 10^{12}\,\mathrm{M}_\odot$. The galaxy stellar masses quoted in this paper should be understood as the sum over all star particles that belong to a galaxy structure identified by ADAPTAHOP. Note that most results are derived from a subsample of galaxies with a mass above $10^9\,\mathrm{M}_\odot$, which corresponds to 300 stellar particles.

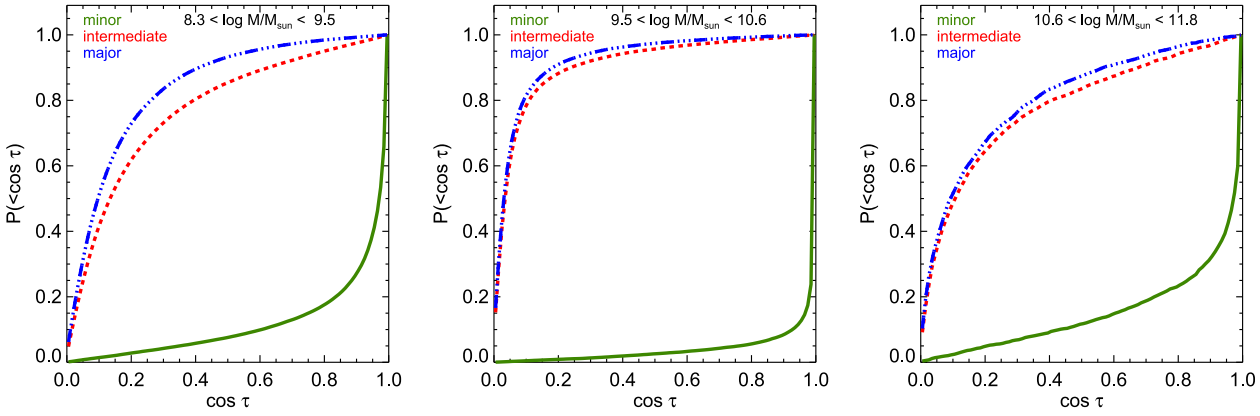


Figure 2. Cumulative PDF of the cosine of the angle, $\cos\tau$, between the galaxy spin and the minor (green solid), intermediate (red dashed) and major (blue dot–dashed) axis of the inertia tensor of the galaxy at $z = 1.2$. Galaxies tend to align their spin with the minor axis. The most massive galaxies, which are more likely dominated by dispersion rather than rotation, show less correlation.

2.2.2 *Spin and shape of galaxies*

To assign a spin to the galaxies, we compute the total angular momentum of the star particles which make up a given galactic structure relative to the particle of maximum density (centre of the galaxy). To identify the latter, we use the smoothed stellar density constructed with the ADAPTAHOP algorithm. We can therefore write the intrinsic angular momentum vector $\boldsymbol{L}$ or spin of a galaxy as

$$\boldsymbol{L} = \sum_{\alpha=1}^{N} m^{(\alpha)} \boldsymbol{x}^{(\alpha)} \times \boldsymbol{v}^{(\alpha)}, \tag{1}$$

where the superscript α denotes the αth stellar particle of mass $m^{(\alpha)}$, position $\boldsymbol{x}^{(\alpha)}$ and velocity $\boldsymbol{v}^{(\alpha)}$ relative to the centre of mass of that galaxy. Likewise, we also measure the (reduced) inertia tensor of a galaxy:

$$I_{ij} = \frac{\sum_{\alpha=1}^{N} m^{(\alpha)} x_i^{(\alpha)} x_j^{(\alpha)}}{\sum_{\alpha=1}^{N} m^{(\alpha)}}. \tag{2}$$

This inertia tensor is then diagonalized to obtain the eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ and the corresponding unit eigenvectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ (respectively minor, intermediate and major axis of the ellipsoid).

Fig. 2 shows the PDF of the angle τ_i between the spin and each of the principle axes $\boldsymbol{u}_i$ of the galaxies:

$$\cos\tau_i = \frac{\boldsymbol{L}\cdot\boldsymbol{u}_i}{|\boldsymbol{L}|}. \tag{3}$$

As expected, galaxies tend to have a spin well aligned with the minor axis ($\boldsymbol{u}_1$) of the inertia tensor, with a mean value $\langle\cos\tau_1\rangle = 0.90$. The correlation is slightly less pronounced for the most massive galaxies, for which the rotation support is weaker compared to velocity dispersion, but it still shows a strong degree of alignment ($\langle\cos\tau_1\rangle = 0.85$).

Owing to the tight alignment between the spin of galaxies and the minor axis of the inertia tensor, we expect that an analysis of galactic spin or inertia tensor orientations will capture the physical mechanisms producing IA equally well. For weak lensing predictions, however, we shall pay attention to the modulus of the ellipticity and not only its direction when turning spins into projected ellipticities. Namely, the projected ellipticity of an axisymmetric discy galaxy depends on the disc thickness (e.g. Joachimi et al. 2013a,b). If we note q_d, the disc flattening or axis ratio, the apparent axis ratio q_p of a projected galaxy along the line of sight aligned with the z-axis, reads

$$q_\mathrm{p} = \frac{|L_z|}{|\boldsymbol{L}|} + q_\mathrm{d}\sqrt{1 - \frac{L_z^2}{|\boldsymbol{L}|^2}}. \tag{4}$$

In principle, we could measure the flattening of the simulated galaxies to infer the projected axis ratios. Unfortunately, the finite $1\,h^{-1}$ kpc resolution of the HORIZON-AGN simulation overestimates the thickness of the disc of low-mass galaxies. To alleviate this problem, we will assume $q_\mathrm{d} = 0$ in the remainder of this work, and thus maximize the moduli of the projected ellipticities. This can be seen as a conservative approach since we effectively maximize the implication of spin IA on apparent alignments of projected ellipticities, either for the correlation between spins themselves (II) or between spins and tidal tensor (GI). We shall come back to the fidelity of our simulations at recovering the eigenvalues of the inertia tensor and not only its eigendirections in a future work.

2.2.3 *Rest-frame intrinsic colours*

In order to ascertain the sensitivity of our measurements to galaxy colours, we compute the absolute AB magnitudes and rest-frame colours of the mock galaxies using single stellar population models from Bruzual & Charlot (2003) adopting a Salpeter initial mass function. Each star particle contributes a flux per frequency that depends on its mass, age and metallicity. The sum of the contribution of all star particles is passed through u, g, r or i filter bands from the Sloan Digital Sky Survey (SDSS). Fluxes are expressed as rest-frame quantities (i.e. that do not take into account the redshifting of spectra) and, for the sake of simplicity, dust extinction is neglected. Once all the star particles have been assigned a flux in each of the colour channels, we build the 2D projected maps for individual galaxies (satellites are excised with the galaxy finder). Summing up the contribution of their stars yields the galaxy luminosity in a given filter band.

Thorough this work, we will investigate the alignment properties of galaxies filtered by their colours. We separate our catalogue in three colour bins (in $u - r$) such that the number of galaxy in each is identical. Our bluest subset corresponds to $u - r < 0.78$ while the reddest one has $u - r > 1.1$. This choice insures a similar noise level in all of our measurements.

3 INTRINSIC ALIGNMENTS IN THE CONTEXT OF COSMIC SHEAR STUDIES

Weak lensing uses the apparent deformation of the shapes of galaxies on the sky to map the gravitational potential or measure its statistical properties like its power spectrum. The projected surface mass density, integrated along the line of sight to distant sources, is often preferred over the potential although they are trivially related by the Poisson equation. The effective convergence κ, which is nothing but the dimensionless projected density, is statistically described by its power spectrum $P_\kappa(\ell)$ as a function of wavenumber ℓ. For a source at comoving distance χ_s, we can write the convergence at an angular position $\boldsymbol{\theta}$ (Bartelmann & Schneider 2001):[3]

$$\kappa(\boldsymbol{\theta}, \chi_s) = \frac{1}{c^2}\int_0^{\chi_s} \mathrm{d}\chi\, \frac{(\chi_s-\chi)\chi}{\chi_s}\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right]\Phi,$$
$$= \frac{3H_0^2\Omega_0}{2c^2}\int_0^{\chi_s} \mathrm{d}\chi\, \frac{(\chi_s-\chi)\chi}{\chi_s}\frac{\delta(\chi\boldsymbol{\theta},\chi)}{a(\chi)}, \quad (5)$$

where χ is the comoving distance, a the expansion factor, δ the density contrast and Φ the three-dimensional gravitational potential are related by the Poisson equation:

$$\Delta\Phi = \frac{3H_0^2\Omega_0}{2a}\delta. \quad (6)$$

This can easily be generalized to a population of sources with a broad redshift distribution (Bartelmann & Schneider 2001).

The relation between Φ and δ can be cast into a relation between the lensing potential ϕ and effective convergence κ, and the effective shear γ_i, which involves the traceless parts of the projected tidal tensor. All these quantities are defined by

$$\phi = \frac{2}{c^2}\int_0^{\chi_s}\mathrm{d}\chi\,\frac{(\chi_s-\chi)\chi}{\chi_s}\Phi, \quad \kappa = \frac{1}{2}\left(\frac{\partial^2\phi}{\partial\theta_1^2}+\frac{\partial^2\phi}{\partial\theta_2^2}\right), \quad (7)$$

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\phi}{\partial\theta_1^2}-\frac{\partial^2\phi}{\partial\theta_2^2}\right), \quad \gamma_2 = \frac{\partial^2\phi}{\partial\theta_1\partial\theta_2}. \quad (8)$$

It is generally suitable to treat the shear in complex notations $\gamma = \gamma_1 + \mathrm{i}\gamma_2$. This quantity is most easily accessible as it captures the amount of anisotropic distortion a light bundle experiences on its way from a distant source to the observer. Therefore, the observed ellipticity of such a source, in the weak lensing regime of small distortions, is directly related to the shear. Indeed, by also defining a complex ellipticity $e = e_1 + \mathrm{i}e_2 = |e|e^{2\mathrm{i}\psi}$, such that $|e| = (1-q)/(1+q)$ and $q = b/a$ is the major (a) to minor (b) axis ratio, we have

$$e = e_s + \gamma, \quad (9)$$

where e is the apparent ellipticity and e_s the intrinsic source ellipticity (the one we would have observed without lensing).

An important statistics of this cosmic shear distortion field is the two-point correlation of projected ellipticities that can formally be split into the following components:

$$\langle e(\vartheta)e(\vartheta+\theta)\rangle_\vartheta = \langle e_s e_s'\rangle + 2\langle e_s\gamma'\rangle + \langle\gamma\gamma'\rangle, \quad (10)$$

where, for compactness, the prime means at an angular distance θ from the first location. The cosmological weak lensing signal is commonly decomposed into the ξ_+ and ξ_- shear correlation functions. Following Schneider et al. (2002), $\xi_\pm$ is given by

$$\xi_\pm(\theta) = \langle\gamma_+\gamma_+\rangle \pm \langle\gamma_\times\gamma_\times\rangle = \frac{1}{2\pi}\int_0^\infty \mathrm{d}\ell\,\ell P_\kappa(\ell)J_{0/4}(\ell\theta),$$

where J_0 and J_4 are the 0th- and 4th-order Bessel functions for ξ_+ and ξ_-, respectively. In this expression, γ_+ (respectively $\gamma_\times$) is the component of the complex shear orientated 0/90° (respectively ±45°) with respect to the line connecting two galaxies separated by a projected distance θ.

The fundamental assumption of weak lensing, which allows to infer shear properties from observed ellipticities, is that, on average, the intrinsic orientation of sources is completely random. The breakdown of this hypothesis yields additional terms to $\langle\gamma\gamma'\rangle$ on the right-hand side of equation (10) that have to be carefully accounted for in observations. The weak lensing signal is therefore contaminated by the two kinds of IA.

(i) The so-called 'II' term $\langle e_s e_s'\rangle$ induced by the intrinsic correlation of the shape of galaxies in the source plane (Croft & Metzler 2000; Heavens et al. 2000; Catelan, Kamionkowski & Blandford 2001). This mostly concerns pairs of galaxies that are at similar redshifts.

(ii) The so-called 'GI' term $\langle e_s\gamma'\rangle$ coming from correlation between the intrinsic ellipticity of a galaxy and the induced ellipticity (or shear) of a source at higher redshift (Hirata & Seljak 2004; Heymans et al. 2006; Joachimi et al. 2011). This non-trivial term is indirectly explained if the shape of galaxies is correlated with the local gravitational tidal field, which also contributes to the shear signal experienced by the far source in a given pair of observed ellipticities.

In this work we propose to measure these two effects in the HORIZON-AGN simulation, Section 4 being devoted to the 'GI' term (essentially captured by spin–tidal field correlations) and Section 5 to the 'II' term (essentially captured by spin–spin correlations).

Before presenting those results, we also give here some guidelines on the way projected correlation functions are worked out in the simulation. In Section 2.2.2, we presented our method for relating three-dimensional galaxy spins $\boldsymbol{L}$ to projected ellipticities in the plane of the sky. It is based on the ansatz that the axis ratio of our galaxies is well approximated by $q = |L_z|/|\boldsymbol{L}|$, where z is the line of sight direction. The orientation of the major axis of the projected ellipse is $\psi = \pi/2 - \arctan(L_y/L_x)$. The projected ellipticities can easily be mapped from Cartesian (1,2) coordinates to the ($+$, $\times$) frame attached to the separation of a given galaxy pair according to the geometric transformation:

$$e_+ = -e_1\cos(2\beta) - e_2\sin(2\beta), \quad (11)$$

$$e_\times = \quad e_1\sin(2\beta) - e_2\cos(2\beta), \quad (12)$$

where β is the angle between the separation and the first Cartesian coordinate.

With those prescriptions, we can estimate the projected correlation functions for a given projected separation θ. For the II component (dropping the subscript s), this reads

$$\xi_+^{\mathrm{II}}(\theta) = \langle e_+e_+' + e_\times e_\times'\rangle. \quad (13)$$

Beyond this 2D measurement, and to limit the dilution of the IA signal with projected angular distances, we will also measure the correlation as a function of the 3D comoving galaxy separation whilst still considering 2D ellipticities as the result of the projection

[3] Simplified to a flat Universe case.

along a specific line of sight. We call this correlation function $\eta(r)$ following the notations of Heymans et al. (2006) and Joachimi et al. (2013b):

$$\eta(r) = \langle e_+(\boldsymbol{x})e_+(\boldsymbol{x}+\boldsymbol{r}) + e_\times(\boldsymbol{x})e_\times(\boldsymbol{x}+\boldsymbol{r})\rangle_x, \quad (14)$$

where $\boldsymbol{r}$ is the 3D galaxy pair separation.

4 SPIN–TIDAL TENSOR CORRELATIONS

In the context of weak lensing surveys, IA can occur through correlations between the shear induced by the gravitational potential in the lens plane, and the intrinsic ellipticity of galaxies in the source plane. Here, we aim to assess the extent to which the tidal tensor and the galactic spins correlate in the HORIZON-AGN simulation described in Section 2. For this purpose, we first measure in Section 4.1 the one-point correlation between the spins and the tidal tensor. Then, Section 4.2 is devoted to the measurements of the two-point correlations between the spins and the tidal tensor as a function of the separation.

To study the correlations between the spin direction and the surrounding gravitational tidal field of the galaxies resolved in the HORIZON-AGN simulation, we measure the components of the 3D (traceless) tidal shear tensor defined as

$$T_{ij} = \partial_{ij}\Phi - \frac{1}{3}\Delta\Phi\,\delta_{ij}, \quad (15)$$

where Φ is the gravitational potential and δ_{ij} the Kronecker δ function. The minor, intermediate and major eigendirections of the tidal tensor T_{ij} are called $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ corresponding to the ordered eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ of the Hessian of the gravitational potential, $\partial_{ij}\Phi$ (with which the tidal tensor shares the eigendirections). In the filamentary regions, $\boldsymbol{e}_1$ gives the direction of the filament (see Fig. 1), while the walls are collapsing along $\boldsymbol{e}_3$ and extend, locally, in the plane spanned by $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ (Pogosyan, Bond & Kofman 1998).

The tidal shear tensor smoothed on scale R_s, $T_{ij} = \partial_{ij}\Phi_{R_s} - \Delta\Phi_{R_s}\,\delta_{ij}/3$, is computed via fast Fourier transform of the density field (including DM, stars, gas and BHs) sampled on a 512^3 Cartesian grid and convolved with a Gaussian filter of comoving scale $R_s = 200\,h^{-1}$ kpc:

$$\partial_{ij}\Phi_{R_s}(\boldsymbol{x}) = \frac{3H_0^2\Omega_0}{2a}\int \mathrm{d}^3\boldsymbol{k}\;\delta(\boldsymbol{k})\frac{k_i k_j}{k^2}W_G(kR_s)\exp(\mathrm{i}\,\boldsymbol{k}\cdot\boldsymbol{x}),$$

where $\delta(\boldsymbol{k})$ is the Fourier transform of the sampled density field and W_G a Gaussian filter.

4.1 One-point cross-correlations

We begin with a measurement of the correlations between the spin and the eigendirections of the tidal tensor at the same spatial position. In practice, we compute the cosine of the angle between the spin of the galaxies and the three eigendirections of the local tidal tensor $\cos\theta = \boldsymbol{L}\cdot\boldsymbol{e}_i/|\boldsymbol{L}|$. The resulting histogram is shown in Fig. 3. The spin is preferentially aligned with the minor eigendirection (i.e. the filaments) in agreement with the spin–filament correlations detected by Dubois et al. (2014) at redshift $z \sim 1.83$. To a lower extent, some alignment is found with the direction of the intermediate axis.

When galaxies are binned in mass (see Fig. 4), it appears that the most massive galaxies tend to have a spin lying in the plane $(\boldsymbol{e}_2, \boldsymbol{e}_3)$ perpendicular to the filaments, while the less massive galaxies have their spin aligned with $\boldsymbol{e}_1$. The transition occurs at stellar masses

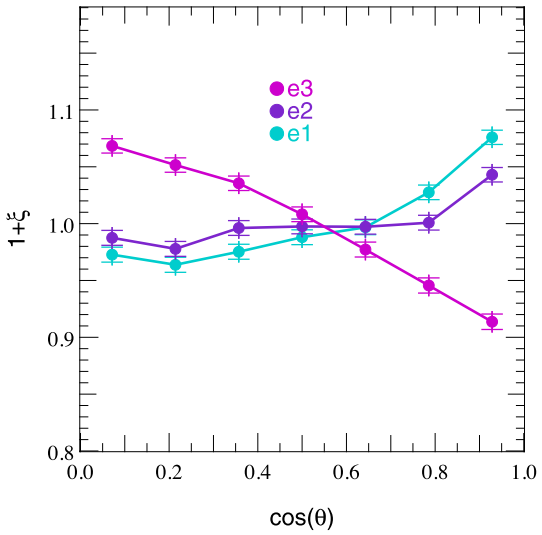


Figure 3. Excess probability of alignment between the spin of galaxies and the minor, intermediate or major axis (respectively cyan, purple and magenta) of the tidal tensor in the HORIZON-AGN simulation. The error bars represent the Poisson noise. The spin of galaxies tends to align with the minor eigendirection. Figs 4 and 5 investigate how this alignment changes with, respectively, galactic mass and colour.

about $4\times10^{10}\,\mathrm{M}_\odot$. We conclude that the spins of *galaxies* are definitely influenced by their surrounding environment differentially with their mass.

Those findings follow very closely what can be found for DM haloes, as detailed in Appendix A where a similar analysis for the alignment of the spin of the DM haloes with the local tidal tensor is carried out. Fig. A1 clearly exhibits the same qualitative correlation as galaxies, namely a transition at a halo mass $\sim5\times10^{11}\,\mathrm{M}_\odot$ from spins aligned with $\boldsymbol{e}_1$, at low mass, to spins oriented in the plane $(\boldsymbol{e}_2, \boldsymbol{e}_3)$ at high mass. This is consistent with previous works based on pure DM simulations (Aragón-Calvo et al. 2007; Hahn et al. 2007a; Paz et al. 2008; Zhang et al. 2009; Codis et al. 2012; Aragón-Calvo 2013; Libeskind et al. 2013, see in particular fig. 3 in Codis et al. 2012).

The colour of galaxies is a quantity more readily accessible to observations. It is therefore of interest to see how different galaxy colours implies different IA as it would provide means on how to leverage this effect. Fig. 5 displays the correlations between galactic spins, and the tidal field for different colours as labelled (see Section 2.2.3 for details about the extraction of galactic colours). The width of the colour bins has been chosen such that there is the same number of objects in each subset of galaxies. On average, the bluest galaxies (defined here by $u - r < 0.78$) are more correlated with the tidal eigendirections than the red galaxies ($u - r > 1.1$ here). This can be easily understood from the fact that red galaxies are typically massive, while blue galaxies are often small-mass galaxies. At that redshift ($z \sim 1.2$), this implies that red galaxies correspond to objects around the transition mass, whereas blue galaxies are mostly aligned with $\boldsymbol{e}_1$. At lower redshift, we expect the population of massive galaxies perpendicular to $\boldsymbol{e}_1$ to increase, so that red galaxies become more correlated. Obviously, we should also keep in mind that applying additional selection cuts on the galaxy samples (mass, luminosity, etc.) would change the level of correlation. For instance, red galaxies above $4\times10^{10}\,\mathrm{M}_\odot$ are more correlated than the whole population of red galaxies.

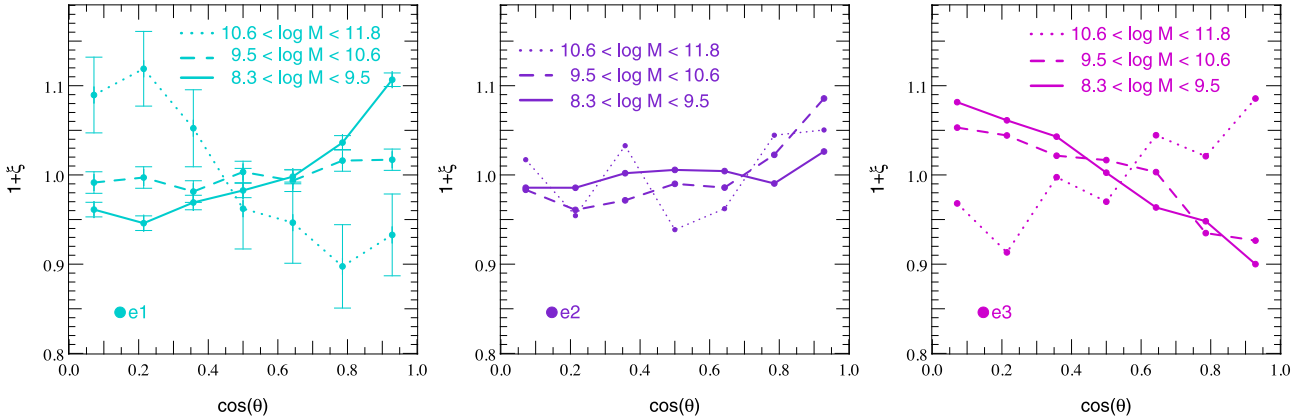


Figure 4. PDF of the cosine of the angle between the spin of galaxies and the minor/intermediate/major axis (from left to right) of the tidal tensor in the HORIZON-AGN simulation when the sample is separated into three different mass bins (solid lines for stellar mass between 2×10^8 and 3×10^9 $M_\odot$, dashed lines for stellar mass between 3×10^9 and 4×10^{10} $M_\odot$ and dotted lines for stellar mass between 4×10^{10} and 6×10^{11} $M_\odot$). The error bars represent the Poisson noise and are only shown for $\boldsymbol{e}_1$ (left-hand panel) since they are the same for $\boldsymbol{e}_2$ (middle panel) and $\boldsymbol{e}_3$ (right-hand panel). The spin of galaxies tends to align with the minor eigendirection at small mass and becomes perpendicular to it at larger mass.

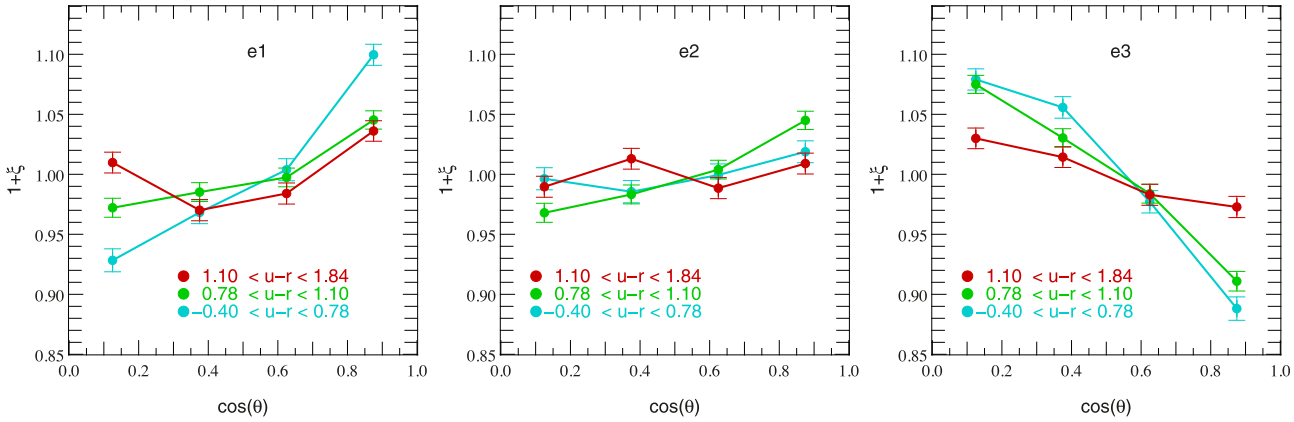


Figure 5. Same as Fig. 4 but for different galaxy colours as labelled, meaning that the left-hand, middle and right-hand panels, respectively, show the PDF of the angle between the $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ directions of the tidal tensor and the galactic spins. The bluer the galaxy, the larger the correlations with the surrounding tidal field. Hence red galaxies are less sensitive to IA.

4.2 Two-point cross-correlations

While the aforementioned measurements have been performed at the same spatial location, it is also of interest in the context of weak lensing studies to quantify how this signal pervades when the separation between galaxies increases. Because the tidal field in the vicinity of a galaxy contributes also to the lensing signal carried by more distant galaxies, it is clear that the spin—-tidal tensor cross-correlation is closely related to the so-called GI term in the weak lensing terminology. To address that question, we measure the correlations between the spins and the eigendirections of the tidal tensor at comoving distance r. In practice, we compute for each pair of galaxies–grid cell (the tidal field being sampled on a 512^3 Cartesian grid) their relative separation and the angle between the spin of the galaxy and the three eigendirections of the tidal tensor in the corresponding grid cell. We finally do a histogram of these quantities. The results are shown in Fig. 6, which displays the PDF of the cosine of the angle between the spins and $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ as a function of the separation, and Fig. 7, which shows on the same plot the mean angle with $\boldsymbol{e}_1$ (cyan), $\boldsymbol{e}_2$ (purple) and $\boldsymbol{e}_3$ (magenta). As expected, the spin and the tidal eigendirections de-correlate with increasing separation. However, whereas the signal vanishes on scales $r > 3\,h^{-1}$ Mpc for the spin to intermediate tidal eigendirection correlation, it persists on distances as large as $\sim 10\,h^{-1}$ Mpc for the minor and major eigendirections of the tidal tensor.

5 SPIN–SPIN AUTOCORRELATIONS

In the previous section, we focused on the correlations between the spins and the tidal tensor eigendirections as it is related to the 'GI term' which is induced by correlations between the ellipticities and the cosmic shear. We will now investigate the second source of IA that comes from the autocorrelations of the intrinsic ellipticities of galaxies. For that purpose, we study first the spin–spin two-point correlation as a function of the galaxy pair separation (Section 5.1), before turning to the projected ellipticity two-point correlation function (Section 5.2).

5.1 3D spin–spin autocorrelations

We begin with the autocorrelation of the direction of the spins as a function of the galaxy pair separation (in other words, the mean angle between the spin of two galaxies separated by a distance r). We select galaxies of different stellar masses: $2 \times 10^8 < M_s < 3 \times 10^9$, $3 \times 10^9 < M_s < 4 \times 10^{10}$ and $M_s > 4 \times 10^{10}$ $M_\odot$ and different

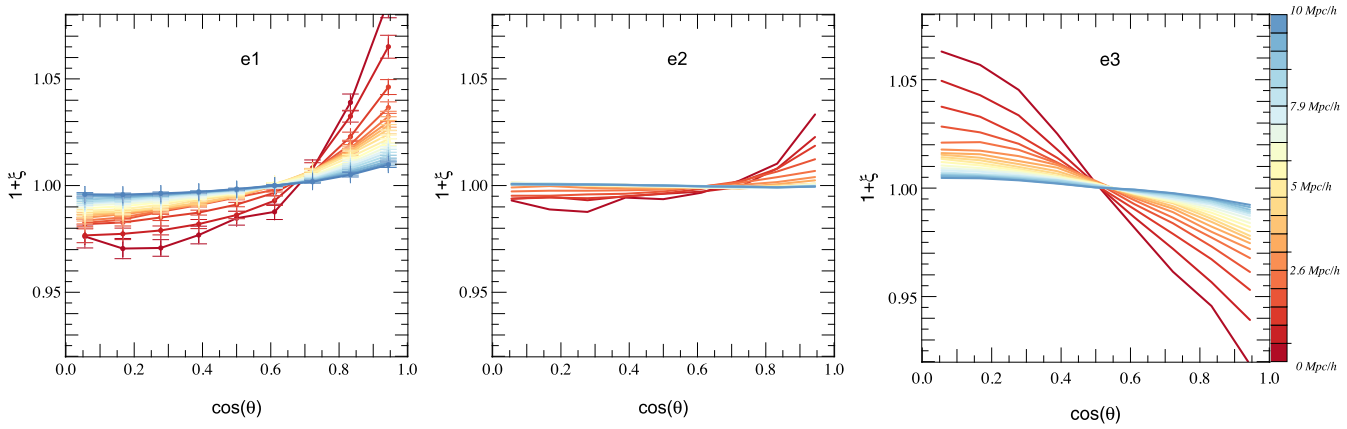


Figure 6. PDF of the cosine of the angle between the spin of galaxies and the tidal tensor (smoothed on 0.4 h^{-1} Mpc) eigendirections $\boldsymbol{e}_1$ (left-hand panel), $\boldsymbol{e}_2$(middle panel) and $\boldsymbol{e}_3$ (right-hand panel) at two locations separated by $r < 10\,h^{-1}$ Mpc (comoving), the separation being colour coded from red to blue. For the sake of readability, the 1σ error on the mean estimating from 32 random resamples is displayed only on the left-hand panel. The spins are more likely to be aligned with $\boldsymbol{e}_1$ and to a lesser extent with $\boldsymbol{e}_2$ at short distance in agreement with the one-point PDF shown in Fig. 3; this signal decreases when the separation increases as expected. This de-correlation is faster for $\boldsymbol{e}_2$ than for the other two eigendirections, going from a few to 10 h^{-1} Mpc.

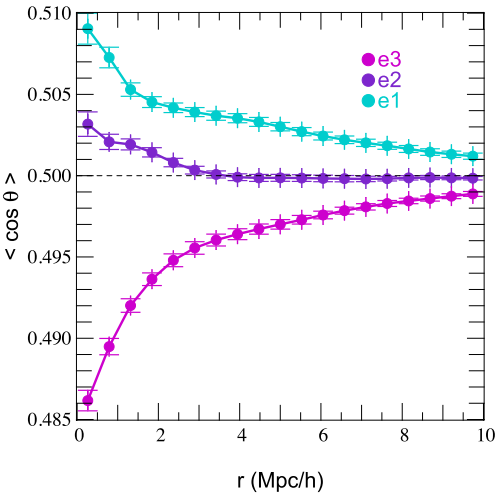


Figure 7. Mean angle between the spin and the tidal eigendirections as a function of the separation. Error bars represent the 1σ error on the mean when cutting the statistics into 32 subsets. The spins are aligned with the minor direction (cyan) and to some extent with the intermediate eigendirection (purple) at small distance and decorrelate on different scales. The alignment with $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$ pervade on scales of about 10 h^{-1} Mpc but is reduced to ∼3 h^{-1} Mpc for the intermediate direction.

colours. For each pair of such galaxies separated by a comoving distance r, we measure the angle between their respective spin, and compute the square of the cosine of this angle, $\cos^2\alpha$ (as the polarity is of no interest for weak lensing), in Fig. 8. Error bars represent the error on the mean $\sqrt{\langle\cos^4\theta\rangle - \langle\cos^2\theta\rangle^2/N}$.

We do not detect any significant spin correlation either among red galaxies or between red and blue galaxies. Conversely we measure a significant spin correlation for blue galaxies out to at least a comoving distance of 10 h^{-1} Mpc. We also see that the correlation amplitude is strong for low- and intermediate-mass galaxies. The signal for the most massive galaxies or red galaxies is compatible with zero correlation at any distance. The importance of grid locking on these correlation is estimated in Section 6 where it is shown that it is not significant.

5.2 2D ellipticity–ellipticity correlations

The two quantities $\xi_+^{\rm II}$ and $\eta(r)$ defined in Section 3 are measured on our synthetic data and shown in Fig. 9. The panels in the top row show our findings for $\eta(r)$ for three populations of galaxies where we choose a direction [here (0.34, 0.06, 0.94) in Cartesian coordinates] in the box different from the grid as the line of sight. Like the previous 3D analysis, there is a striking difference of behaviour between red and blue galaxies, the latter showing a strong correlation signal for the statistics of η with a typical amplitude of $\sim 3 \times 10^{-3}$ between 1 and 5 h^{-1} Mpc. On the other hand, the correlation for red galaxies is compatible with zero. We can therefore anticipate that blue galaxies at redshift ∼1.2 should be affected by IA, leaving the possibility of a substantial contamination of the weak lensing signal. The bottom panels of Fig. 9 show the amplitude of the correlation function $\xi_+^{\rm II}(\theta)$ as a function of the angular galaxy pairs separation.[4] It is significant and comparable to the cosmic shear amplitude all the way to ∼13 arcmin.

Let us now compare our findings to the recent study of Joachimi et al. (2013b), which is based on a semi-analytical model of galaxy formation. At first glance, we draw opposite conclusions, as we measure a strong level of IA for blue galaxies and no noticeable signal for red galaxies, whereas Joachimi et al. predict a strong alignment of red galaxies ($\eta \sim 10^{-2}$ at $z = 1.5$ and $r = 1\,h^{-1}$ Mpc) and a low level of alignment of blue galaxies ($\lesssim 3 \times 10^{-4}$ at $r = 1\,h^{-1}$ Mpc, compatible with zero correlation). Yet, several important differences should be emphasized.

(i) All our galaxies are assumed to be thin discs with a value $q_{\rm d} = 0$ in equation (4) whereas Joachimi et al. explore two values $q_{\rm d} = 0.1$ and 0.25 for late-type galaxies. Recall that our choice of $q_{\rm d} = 0$ will tend to maximize the amplitude of projected ellipticities. We repeated the measurement of $\xi_+^{\rm II}$ assuming a value of $q_{\rm d} = 0.25$ which assumes quite a strong thickening of the disc. The net result shown in Fig. 10 is to decrease the normalization of $\xi_+^{\rm II}$ for blue galaxies by a factor of ∼3, to a typical amplitude of $\xi_+^{\rm II} \simeq 2 \times 10^{-4}$

[4] Note that a $1\,h^{-1}$ Mpc comoving transverse distance corresponds to an angular size of 1.3 arcmin at this redshift.

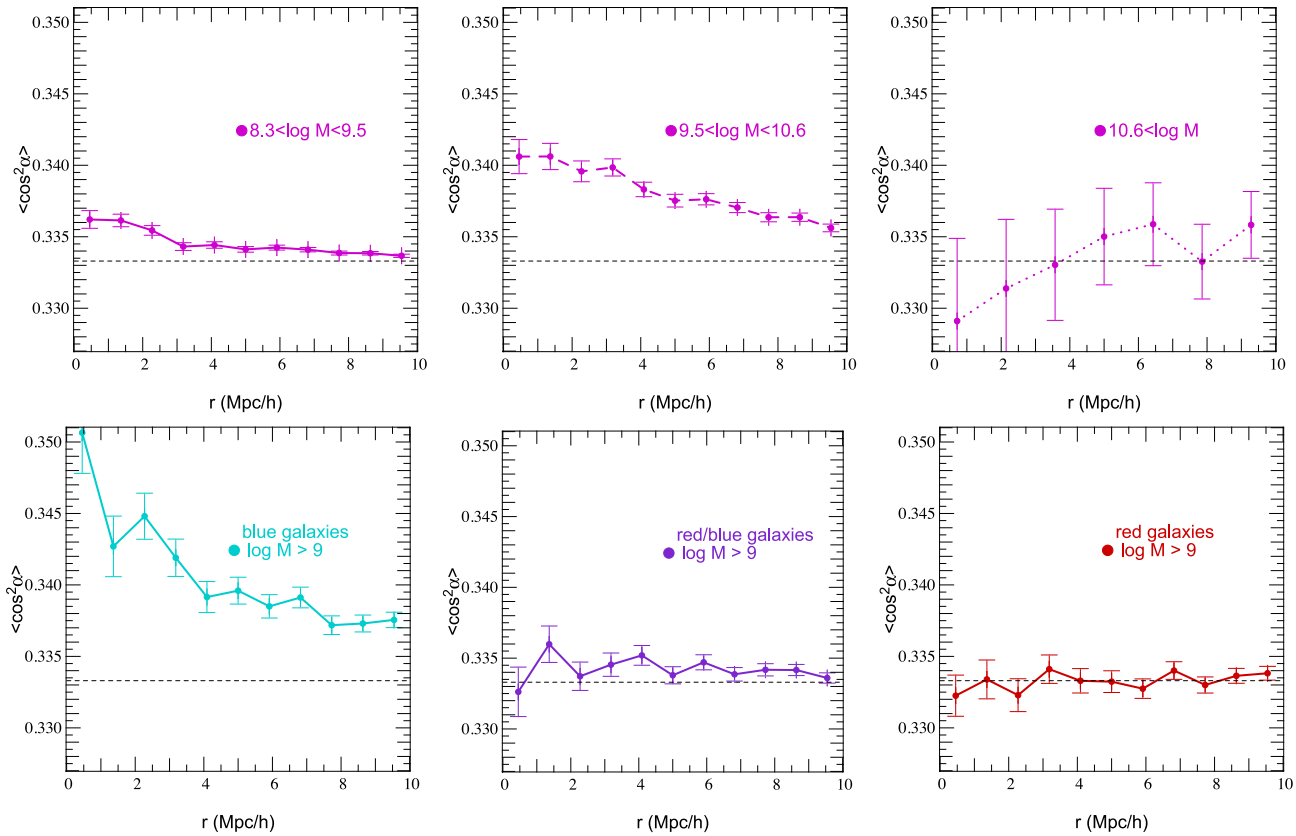


Figure 8. 3D spin–spin two-point correlation function of galaxy as a function of the comoving separation for a range of stellar masses: $2 \times 10^8 < M_s < 3 \times 10^9\,M_\odot$ (top left-hand panel), $3 \times 10^9 < M_s < 4 \times 10^{10}\,M_\odot$ (top centre panel) and $M_s > 4 \times 10^{10}\,M_\odot$ (top right-hand panel), or colours: blue galaxies, i.e. $u - r < 0.78$ (bottom left-hand panel), red galaxies, i.e. $u - r > 1.1$ (bottom right-hand panel) and the cross-correlations between blue and red galaxies (bottom centre panel). Note that for blue and red galaxies we also apply a mass cut below $10^9\,M_\odot$ (i.e. 300 star particles). Error bars represent the error on the mean. Here we choose to display the mean square cosine between two spins (separated by the comoving distance r) as the polarity is not relevant to weak lensing studies. For a uniform random distribution, the expectation is 1/3 (dashed line). Blue galaxies and small-to-middle mass galaxies are indeed correlated on the scale of the typical size of filaments, $\gtrsim 10\,h^{-1}$ Mpc, whereas red and high-mass galaxies do not show a significant correlation.

between 1 and 5 h^{-1} Mpc at $z \sim 1.2$ (but for intermediate-mass galaxies, the amplitude of the correlation remains qualitatively the same).

(ii) Our distinction of red and blue galaxies may not correspond exactly to their early- and late-type classification. For instance, our red galaxies sample is dominated by low-mass red satellite instead of massive red central galaxies. But this effect cannot fully explain the difference.

(iii) To assign a spin to their late-type galaxies (somehow equivalent to our blue sample), Joachimi et al. used the results of Bett (2012) for the alignment between the DM halo spin orientation and the galaxy spin at redshift zero. These were obtained with several hydrodynamical zoom simulations in a $\sim$12.5 or $\sim 20\,h^{-1}$ Mpc size volume. As shown in Fig. A2, we find at redshift $z \sim 1.2$ a somewhat weaker alignment between galaxy and halo spins than Bett (2012). Naively, this fact should produce weaker two-point galaxy spin–spin correlation than the Joachimi et al. results. However, Dubois et al. (2014) showed that the spin of galaxies are in fact as correlated with the large-scale filaments as their DM host halo. This is a consequence of cold flows that advect efficiently the cosmic angular momentum all the way to galaxies at the centre of dark haloes (Kimm et al. 2011; Pichon et al. 2011; Danovich et al. 2012; Stewart et al. 2013). Therefore semi-analytic models, which chain the de-correlation between the large-scale structure and the halo, and that between the halo and the disc galaxies, will most certainly underestimate the correlation of these galaxies with the large-scale structure.

(iv) Conversely, the most likely explanation for the discrepancy of the red population is twofold. First, the red galaxies here are objects around the transition mass. Hence, part of them have their spin aligned with the surrounding filament (below the transition mass), while the others have their spin perpendicular to it (above the transition mass). Therefore, on average no correlation is detected among that sample. Nevertheless, for larger mass and lower redshift, this population is expected to become more strongly perpendicular to the filaments and thus more correlated. Again, one should bear in mind that any selection (mass, colour, luminosity, etc.) may bias significantly the two-point ellipticity correlations. Second, spins alone may not fully capture the shape of non-rotating, mostly triaxial early-type galaxies. We defer for future work a thorough analysis of ellipticity alignments with the inertia tensor of galaxies as a proxy for ellipticity.

Note finally that the coherence of the large-scale structure is expected to be higher with increasing redshift. We may therefore speculate that, at later times, the intrinsic correlation of spins will decrease (see also Lee et al. 2008; Joachimi et al. 2013b).

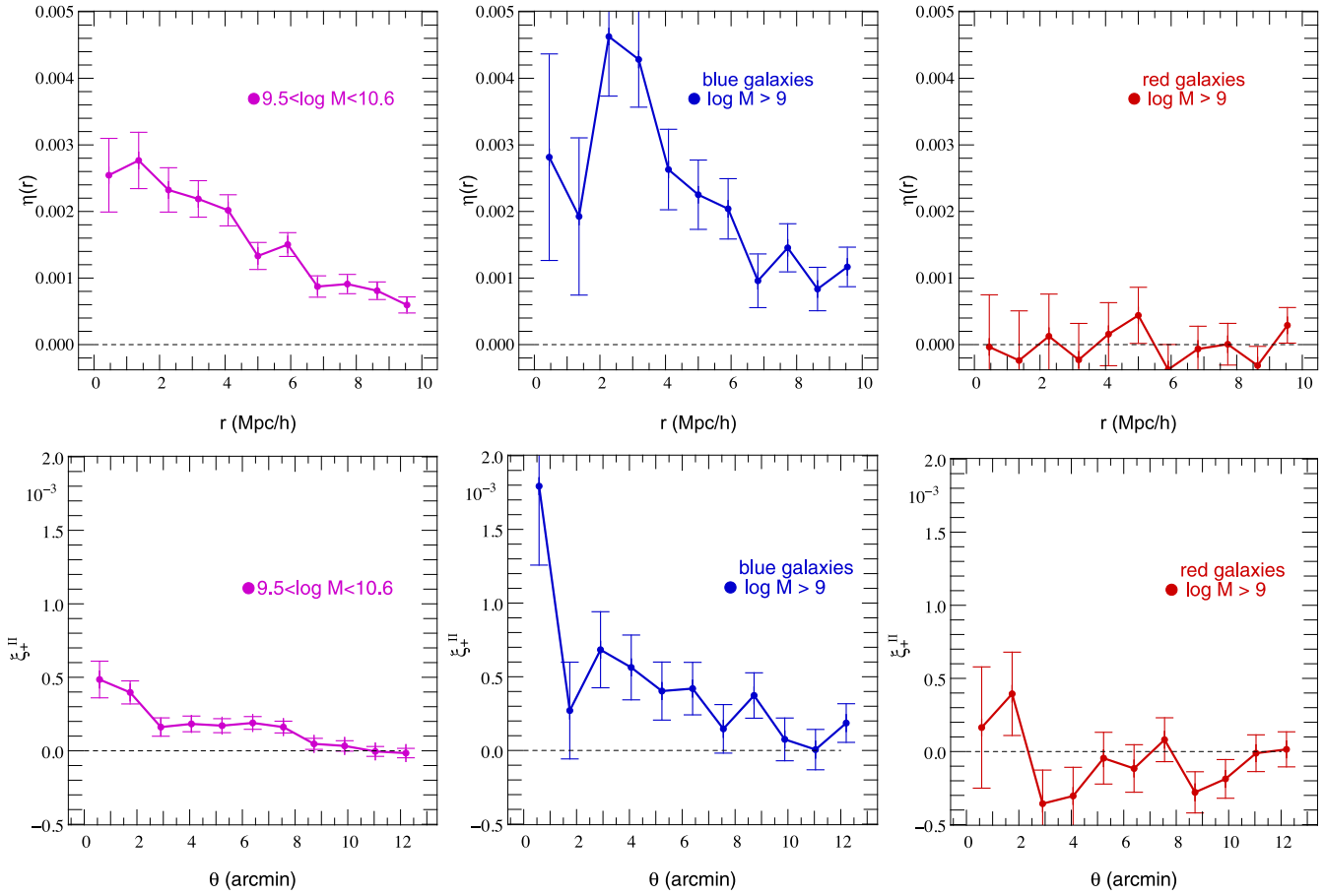


Figure 9. Correlation functions of the projected ellipticity $\eta(r)$ for the ∼58 000 middle-mass (left-hand panels), ∼25 000 blue (middle panels) and ∼25 000 red (right-hand panels) galaxies as a function of the comoving 3D separation r, in the top row, and as a function of the (projected) angular separation θ in the bottom row; this latter quantity being closer to observations, we call it $\xi_+^{\rm II}$. Note the change in scale from one row to the other and the additional mass cut below $10^9\,{\rm M}_\odot$ (i.e. 300 star particles).

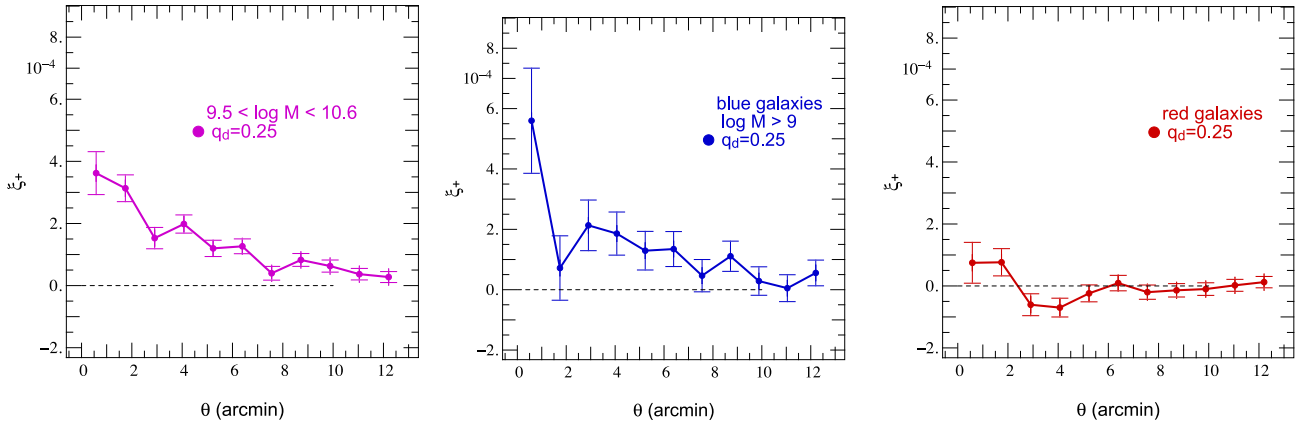


Figure 10. Same as bottom panels of Fig. 9 for thicker discs $q_{\rm d} = 0.25$.

6 GRID LOCKING EFFECT ON SPIN ALIGNMENTS

Grid locking effects are a concern for spin–spin correlation functions when using grid-based codes such as AMR (see also Dubois et al. 2014 for details about how spins are correlated with the Cartesian grid). The alignments of the spins with the grid are shown in Fig 11. To investigate this effect on the two-point statistics studied in this work, we compute the same statistics (spin–spin correlation functions and its 2D counterpart, $\xi_+^{\rm II}$) after a random permutation of the spins in the box. This allows us to keep the same one-point distribution of spins on the sphere, including the corresponding level of grid locking, but to remove the physical two-point correlations (the box size being much larger than the typical correlation length). Thus any signal in the data, relative to the random permutation of spins, should be physical and not induced by grid locking. Fig. 12

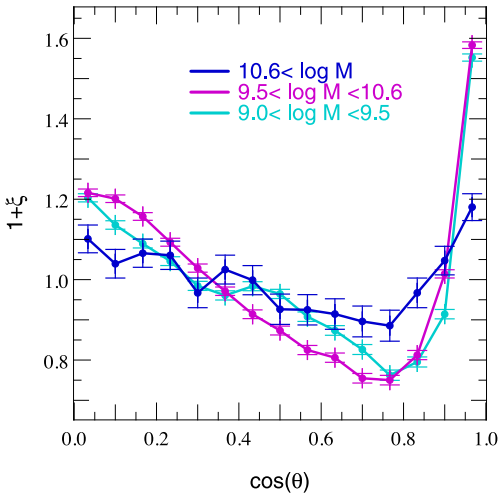


Figure 11. Excess probability of alignment, ξ, between the spins and the grid axes for different bins of mass. The most massive galaxies (blue solid line) are the least prone to grid locking. Note that the displayed plot is the mean on the PDF obtained for the x, y and z axes.

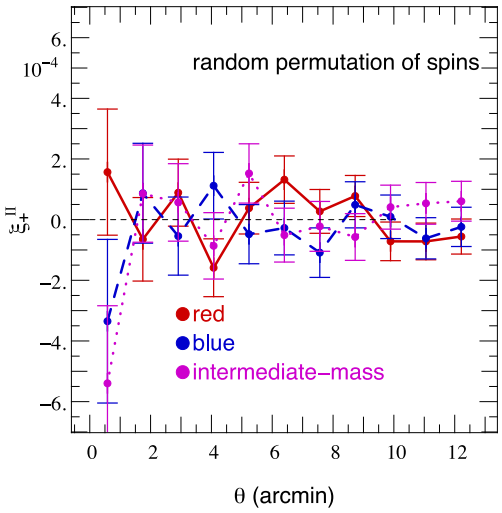


Figure 12. $\xi_+^{\rm II}$, the two-point projected ellipticity correlation function for intermediate-mass, blue and red galaxies when spins are randomly swapped in the box. The three panels are statistically compatible with non-correlation, meaning that grid locking does not introduce significant ellipticity correlations. This allows us to assume that the correlations detected in Fig. 9 are physical and not due to grid locking effects.

shows the result for different populations of galaxies (blue, red and intermediate mass). It appears that all those correlations are consistent with $\xi_+^{\rm II} = 0$, implying that the results presented in the main text (e.g. Fig. 9) are not significantly biased by grid locking. The same conclusion holds for spin–spin correlations and $\eta(r)$.

To be more precise, let us assume that the spin of a galaxy in the simulation is the superposition of the 'real' spin plus a numerical contribution coming from grid locking $\boldsymbol{s} = \boldsymbol{s}_{\rm s} + \boldsymbol{s}_{\rm gl}$ where we neglect normalizations (all spins are of norm 1 here). Then the two-point function of the measured spin of galaxies reads

$$\langle \boldsymbol{s}(\boldsymbol{x})\boldsymbol{s}(\boldsymbol{x}+\boldsymbol{r})\rangle = \langle \boldsymbol{s}_{\rm s}(\boldsymbol{x})\boldsymbol{s}_{\rm s}(\boldsymbol{x}+\boldsymbol{r})\rangle + 2\left\langle \boldsymbol{s}_{\rm s}(\boldsymbol{x})\boldsymbol{s}_{\rm gl}(\boldsymbol{x}+\boldsymbol{r})\right\rangle + \left\langle \boldsymbol{s}_{\rm gl}(\boldsymbol{x})\boldsymbol{s}_{\rm gl}(\boldsymbol{x}+\boldsymbol{r})\right\rangle . \quad (16)$$

Let us also assume that the spin contributions coming from grid locking do not depend on spatial location. Then $\left\langle \boldsymbol{s}_{\rm s}(\boldsymbol{x})\boldsymbol{s}_{\rm gl}(\boldsymbol{x}+\boldsymbol{r})\right\rangle$ and $\left\langle \boldsymbol{s}_{\rm gl}(\boldsymbol{x})\boldsymbol{s}_{\rm gl}(\boldsymbol{x}+\boldsymbol{r})\right\rangle$ do not depend on $\boldsymbol{r}$, i.e. those terms are constant. Therefore one can write

$$\langle \boldsymbol{s}(\boldsymbol{x})\boldsymbol{s}(\boldsymbol{x}+\boldsymbol{r})\rangle = \langle \boldsymbol{s}_{\rm s}(\boldsymbol{x})\boldsymbol{s}_{\rm s}(\boldsymbol{x}+\boldsymbol{r})\rangle + C. \quad (17)$$

If one permutes all the spins in the box the physical two-point correlation of spins $\langle \boldsymbol{s}_{\rm s}(\boldsymbol{x})\boldsymbol{s}_{\rm s}(\boldsymbol{x}+\boldsymbol{r})\rangle$ goes to zero if the box is large enough and C remains C as it does not depend on the spatial coordinates. We therefore measured this term in the simulation and found that $|C| \lesssim 10^{-4}$, which is below the statistical uncertainties currently plaguing our physical measurements. If grid locking does not depend on spatial location, we can conclude that the measured spin–spin correlation function is not significantly biased.

Another test that one can perform is to restrict the analysis to the galaxies that are less grid locked. We therefore measured the spin–spin correlation function for the most massive galaxies (that were shown to be the least sensitive to grid locking in Fig. 11). The result is displayed in Fig. 13. As the signal remains qualitatively the same (the amplitude is even larger), we conclude that grid locking cannot be the main source of spin correlations measured in the simulation. Note that this test does not make any assumption on the physical origin of grid locking.

The only way to go beyond the tests proposed in this section would be to compare our results with similar cosmological hydrodynamics simulations performed using a technique which does not suffer from grid locking, but this is clearly beyond the scope of this paper.

7 SUMMARY AND CONCLUSION

Using the HORIZON-AGN simulation, we have shown that low-mass galaxies tend to have their spin aligned with the local tidal field minor eigendirection (the filamentary direction), whereas more massive galaxies ($M_{\rm s} > 4 \times 10^{10}\,{\rm M}_\odot$ at redshift one) have their spin mostly aligned with the major tidal eigendirection (i.e. perpendicular to walls and filaments). The corresponding two-point correlation decreases with the comoving separation out to scales as large as $\sim 10\,h^{-1}$ Mpc, with a faster de-correlation for the spin-to-the-intermediate-axis direction ($\sim 3\,h^{-1}$ Mpc), a result consistent with Lagrangian theory in this context (Codis et al., in preparation). Those results depend on the properties of galaxies, in particular on their mass and intrinsic colour. For instance, it was clearly found that at $z \simeq 1.2$ in the HORIZON-AGN simulation, blue galaxies and intermediate-mass galaxies are significantly correlated with the gravitational tidal field whereas red and massive galaxies do not show any correlations. We reach identical conclusions when studying the alignment of galaxy spins with one another, namely that the spins of galaxies are also correlated on similar scales ($\sim 10\ h^{-1}$ Mpc) and are similarly colour and mass dependent.

We have also investigated how spin–spin correlations project into weak lensing observables like the shear correlation function ξ_+, these correlations being cast into the so-called II contributions to IA. As in 3D, a $\xi_+^{\rm II}$ correlation at a level of a few 10^{-4} is found for blue and intermediate-mass galaxies out to separations of $\gtrsim 10$ arcmin for sources at redshift ~ 1.2. The results for blue galaxies are in broad agreement with the recent work of Joachimi et al. (2013b), who combine observational results on IA from the Cosmic Evolution Survey (COSMOS) and predictions from semi-analytical models applied to DM-only simulations. The effect of grid locking on the two-point functions was shown to be subdominant. However,

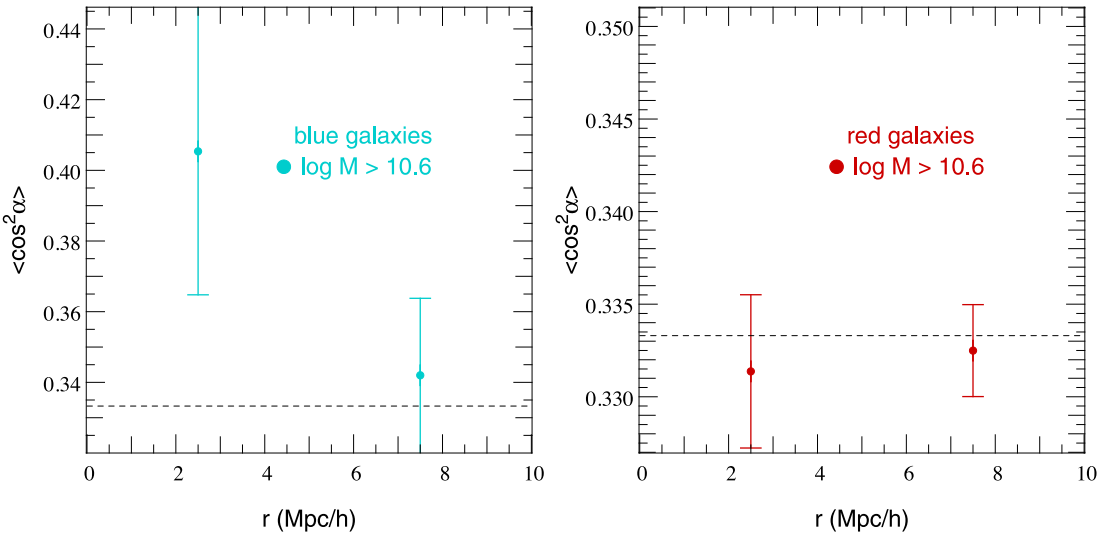


Figure 13. 3D spin–spin two-point correlation function of galaxy as a function of the comoving separation for blue galaxies, i.e. $u - r < 0.78$ (left-hand panel) and red galaxies, i.e. $u - r > 1.1$ (right-hand panel) with a mass cut below $4 \times 10^{10}\,M_{\odot}$. Same qualitative results as in Fig. 8 are found. Hence, grid locking should not be the dominant source of the observed spin correlations.

to go beyond this qualitative statement and accurately quantify the effect of grid locking, one would need to perform a similar simulation using a numerical technique insensitive to this specific systematic error, such as smoothed particle hydrodynamics (SPH) or unstructured mesh.

Presently, the 'spin-gives-ellipticity' prescription allows one to quantify the new insights that large volume hydrodynamical cosmological simulations bring to the issue of IA. For instance, the large-scale coherence of gas motions advected all the way to the centre of galaxies through cold flows regardless of the DM behaviour can uniquely be captured by such simulations (Kimm et al. 2011). Large-scale dynamics imprint their coherence and morphology (filaments, walls, voids) on to the spin of galaxies. This complex topology is likely to have an even more prominent impact on higher order statistics beyond the shear two-point correlation function. Attempts to capture such effects with simple halo occupancy distribution prescriptions may therefore fail at high redshift ($z \gtrsim 0.8$), which is the place where galaxies carry more cosmological lensing signal and is also, to large extent, the population of sources targeted by future surveys like *Euclid* or LSST. The challenge for simulations is to cover large cosmological volumes while preserving a sufficient resolution so that baryonic physics (star formation, feedback processes, etc.) is correctly treated.

When the HORIZON-AGN simulation reaches redshift zero, we will be in a good position to compare our findings with existing observations. In order to get a good match for massive red galaxies, we will certainly adopt a different ansatz for our recipe – currently based on a thin disc approximation – and use directly the resolved shape of massive galaxies as a proxy for the projected ellipticities. However, this concerns only a small fraction of the galaxies that made up the typical weak lensing catalogues of background sources. Once the HORIZON-AGN light-cone is completed, we will estimate more realistic galactic shapes, taking full account of the spectral energy distribution of young and old stars (giving a non-trivial weight to the relative contribution of the disc and the bulge) into a well chosen rest-frame filter (e.g. using the broad *Euclid* VIS band), and more precisely mimicking observational selection effects (i.e. a flux-limited sample of background sources will capture specific populations of sources at a given redshift). This will allow us to quantify the amount of contamination from IA expected in real surveys, and possibly mitigate their nuisance by selecting galaxies that are less prone to IA based on colour.

ACKNOWLEDGEMENTS

This work has made use of the HPC resources of CINES (Jade supercomputer) under the allocation 2013047012 and 2014047012 made by GENCI. Part of the analysis was performed on the DiRAC facility jointly funded by STFC and the Large Facilities Capital Fund of BIS. This work is partially supported by the Spin(e) grants ANR-13-BS05-0005 of the French *Agence Nationale de la Recherche* and by the ILP LABEX (under reference ANR-10-LABX-63 and ANR-11-IDEX-0004-02). The research of YD has been supported at IAP by ERC project 267117 (DARK) hosted by Université Pierre et Marie Curie – Paris 6. We thank S. Rouberol for running smoothly the HORIZON cluster for us, and D. Munro for freely distributing his YORICK programming language and OPENGL interface (available at http://yorick.sourceforge.net/). SC and CP thank Lena for her hospitality during the course of this work. VD would like to thank the Institut d'Astrophysique de Paris for hospitality during the completion of this work, and acknowledges support by the Swiss National Science Foundation. JD and ASs research is supported by funding from Adrian Beecroft, the Oxford Martin School and the STFC.

REFERENCES

Aragón-Calvo M. A., 2013, preprint (arXiv:e-prints)
Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007, ApJ, 655, L5
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bailin J., Steinmetz M., 2005, ApJ, 627, 647
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bernstein G. M., Norberg P., 2002, AJ, 124, 733
Bett P., 2012, MNRAS, 420, 3303
Bett P., Eke V., Frenk C. S., Jenkins A., Okamoto T., 2010, MNRAS, 404, 1137
Blazek J., Mandelbaum R., Seljak U., Nakajima R., 2012, J. Cosmol. Astropart. Phys., 5, 41
Booth C. M., Schaye J., 2009, MNRAS, 398, 53

Bridle S., King L., 2007, New J. Phys., 9, 444
Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002, MNRAS, 333, 501
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Catelan P., Kamionkowski M., Blandford R. D., 2001, MNRAS, 320, L7
Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320
Croft R. A. C., Metzler C. A., 2000, ApJ, 545, 561
Danovich M., Dekel A., Hahn O., Teyssier R., 2012, MNRAS, 422, 1732
Deason A. J. et al., 2011, MNRAS, 415, 2607
Dubois Y., Teyssier R., 2008, A&A, 477, 79
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, MNRAS, 420, 2662
Dubois Y. et al., 2014, MNRAS, 444, 1453
Forero-Romero J. E., Contreras S., Padilla N., 2014, MNRAS, 443, 1090
Greggio L., Renzini A., 1983, A&A, 118, 217
Haardt F., Madau P., 1996, ApJ, 461, 20
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007a, MNRAS, 375, 489
Hahn O., Carollo C. M., Porciani C., Dekel A., 2007b, MNRAS, 381, 41
Hahn O., Teyssier R., Carollo C. M., 2010, MNRAS, 405, 274
Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649
Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, MNRAS, 347, 895
Heymans C., White M., Heavens A., Vale C., van Waerbeke L., 2006, MNRAS, 371, 750
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Hirata C. M. et al., 2004, MNRAS, 353, 529
Hirata C. M., Mandelbaum R., Ishak M., Seljak U., Nichol R., Pimbblet K. A., Ross N. P., Wake D., 2007, MNRAS, 381, 1197
Joachimi B., Bridle S. L., 2010, A&A, 523, A1
Joachimi B., Schneider P., 2008, A&A, 488, 829
Joachimi B., Schneider P., 2010, A&A, 517, A4
Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, A&A, 527, A26
Joachimi B., Semboloni E., Bett P. E., Hartlap J., Hilbert S., Hoekstra H., Schneider P., Schrabback T., 2013a, MNRAS, 431, 477
Joachimi B., Semboloni E., Hilbert S., Bett P. E., Hartlap J., Hoekstra H., Schneider P., 2013b, MNRAS, 436, 819
Kennicutt R. C., Jr, 1998, ApJ, 498, 541
Kimm T., Devriendt J., Slyz A., Pichon C., Kassin S. A., Dubois Y., 2011, preprint (arXiv:e-prints)
Kirk D., Bridle S., Schneider M., 2010, MNRAS, 408, 1502
Komatsu E. et al., 2011, ApJS, 192, 18
Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304
Laigle C. et al., 2015, MNRAS, 446, 2744
Laureijs R. et al., 2011, preprint (arXiv:e-prints)
Lee J., Pen U.-L., 2002, ApJ, 567, L111
Lee J., Springel V., Pen U.-L., Lemson G., 2008, MNRAS, 389, 1266
Lee B. et al., 2013, ApJ, 774, 47
Leitherer C. et al., 1999, ApJS, 123, 3
Leitherer C., Ortiz Otálvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, ApJS, 189, 309
Libeskind N. I., Hoffman Y., Forero-Romero J., Gottlöber S., Knebe A., Steinmetz M., Klypin A., 2013, MNRAS, 428, 2489
Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, MNRAS, 367, 611
Mandelbaum R. et al., 2011, MNRAS, 410, 844
Nelson D., Vogelsberger M., Genel S., Sijacki D., Kereš D., Springel V., Hernquist L., 2013, MNRAS, 429, 3353
Paz D. J., Stasyszyn F., Padilla N. D., 2008, MNRAS, 389, 1127
Pichon C., Pogosyan D., Kimm T., Slyz A., Devriendt J., Dubois Y., 2011, MNRAS, 418, 2493
Pichon C., Codis S., Pogosyan D., Dubois Y., Desjacques V., Devriendt J., 2014, arXiv e-prints
Pogosyan D., Bond J. R., Kofman L., 1998, J. R. Astron. Soc. Canada, 92, 313
Porciani C., Dekel A., Hoffman Y., 2002, MNRAS, 332, 325
Rasera Y., Teyssier R., 2006, A&A, 445, 1
Salpeter E. E., 1955, ApJ, 121, 161
Scannapieco C. et al., 2012, MNRAS, 423, 1726
Schneider M. D., Bridle S., 2010, MNRAS, 402, 2127
Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002, A&A, 396, 1
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Sousbie T., Pichon C., Colombi S., Pogosyan D., 2008, MNRAS, 383, 1655
Stewart K. R., Brooks A. M., Bullock J. S., Maller A. H., Diemand J., Wadsley J., Moustakas L. A., 2013, ApJ, 769, 74
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Tenneti A., Mandelbaum R., Di Matteo T., Feng Y., Khandai N., 2014, MNRAS, 441, 470
Teyssier R., 2002, A&A, 385, 337
Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747

APPENDIX A: DARK MATTER HALOES

Fig. A1 displays the PDF of the cosine of the angle between the tidal eigendirections and the spin of DM haloes of different masses. The same qualitative behaviour as for galaxies is detected: the less massive haloes have a spin preferentially aligned with $\boldsymbol{e}_1$ (somehow

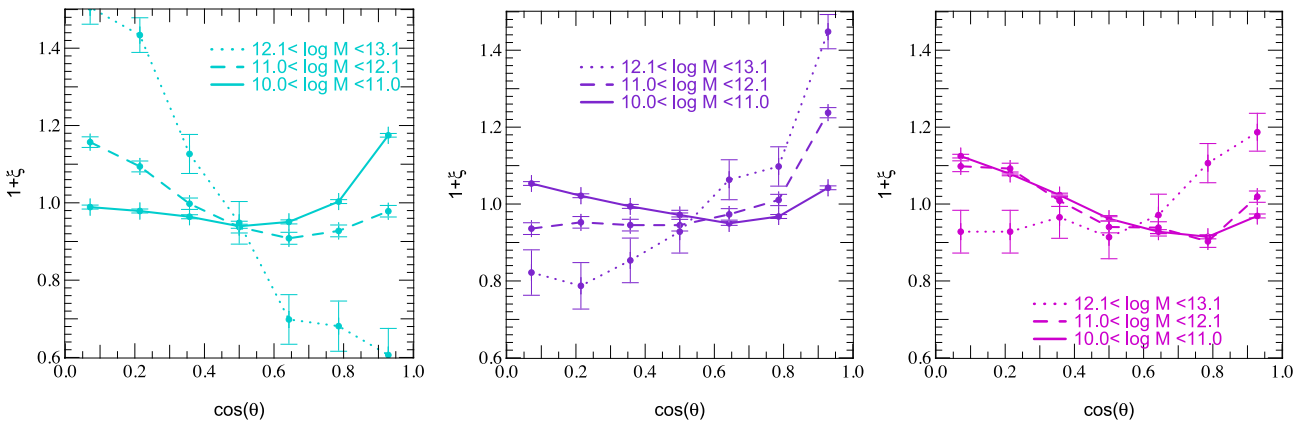


Figure A1. Same as Fig. 3 for DM haloes in the HORIZON-AGN simulation. The three different axes of the tidal tensor are colour coded from cyan ($\boldsymbol{e}_1$) to magenta ($\boldsymbol{e}_3$) through purple ($\boldsymbol{e}_2$). Different mass bins are coded from solid (10^{10}–10^{11} $M_\odot$) to dotted (10^{12}–10^{13} $M_\odot$) through dashed lines (10^{11}–10^{12} $M_\odot$). A transition is detected: the spin of high-mass haloes tends to be aligned with the intermediate (centre purple panel) and with less probability major (right-hand magenta panel) principal axis, whereas the spin of low-mass haloes is more likely to point along the minor axis (left-hand cyan panel). Those findings are consistent with previous studies of DM-only simulations.

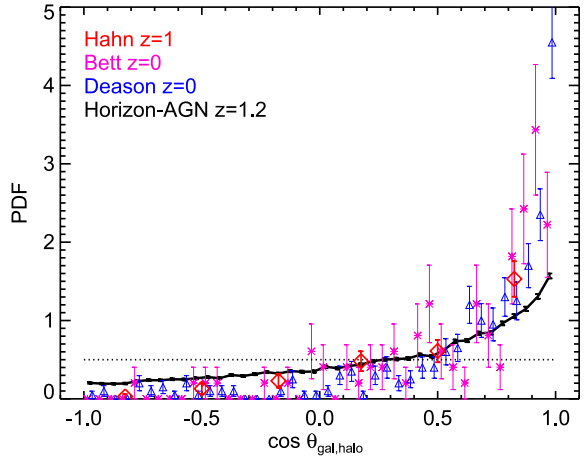


Figure A2. PDF of the cosine of the angle between the spin of the DM halo and that of the galaxy in our simulation at $z = 1.2$ (solid) and from Hahn et al. (2010) at $z = 1$ (red diamonds), from Bett et al. (2010) (pink stars) and from Deason et al. (2011) (blue triangles) both at $z = 0$. The dotted line refers to a uniform isotropic distribution. Error bars are the 1σ standard deviation for Poisson statistics.

the filaments) while higher masses have their spin perpendicular to it (in agreement with what previous studies found in DM-only simulations).

Fig. A2 shows the PDF of the cosine of the angle between the spin of the DM halo and the spin of the galaxy for ∼32 000 pairs at $z = 1.2$ for haloes with virial mass $M_{\rm h} > 10^{11}\,{\rm M}_{\odot}$. We compare our result to that of Hahn et al. (2010) ($z = 1$, $M_{\rm h} > 10^{11}\,{\rm M}_{\odot}$, 89 pairs), Bett et al. (2010) ($z = 0$, $M_{\rm h} > 5 \times 10^{9}\,{\rm M}_{\odot}$, 99 pairs) and Deason et al. (2011) ($z = 0, 5 \times 10^{11} < M_{\rm h} < 5 \times 10^{12}\,{\rm M}_{\odot}$, 431 pairs). It shows that galaxies are slightly less aligned with their host DM halo in our HORIZON-AGN simulation than what is found by other works, and used in the semi-analytic model of Joachimi et al. (2013b) (based on Bett et al. 2010; Deason et al. 2011), even though all results are compatible within 2σ error bars. This slight difference might originate from AMR (HORIZON-AGN; Hahn et al. 2010) versus SPH (Bett et al. 2010; Deason et al. 2011), the persistence of cold flows (Nelson et al. 2013), disc versus elliptical galaxies (Scannapieco et al. 2012), representativity of their resimulations, etc. Note that both AMR runs (HORIZON-AGN and that of Hahn et al. 2010 performed with the RAMSES code as well) at the same redshift are in better agreement.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed September 16, 2015 (MN LATEX style file v2.2)

Intrinsic alignments of galaxies in the Horizon-AGN cosmological hydrodynamical simulation

N. Chisari[1], S. Codis[2], C. Laigle[2], Y. Dubois[2], C. Pichon[2,3], J. Devriendt[1], A. Slyz[1], L. Miller[1], R. Gavazzi[2] and K. Benabed[2]

[1] *Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH,United Kingdom.*
[2] *Institut d'Astrophysique de Paris, CNRS & UPMC, UMR 7095, 98 bis Boulevard Arago, 75014, Paris, France.*
[3] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, United Kingdom.*

Accepted September 15, 2015. Received July 31, 2015; in current form September 16, 2015

ABSTRACT

The intrinsic alignments of galaxies are recognised as a contaminant to weak gravitational lensing measurements. In this work, we study the alignment of galaxy shapes and spins at low redshift ($z \sim 0.5$) in Horizon-AGN, an adaptive-mesh-refinement hydrodynamical cosmological simulation box of $100\,h^{-1}\,$Mpc a side with AGN feedback implementation. We find that spheroidal galaxies in the simulation show a tendency to be aligned radially towards over-densities in the dark matter density field and other spheroidals. This trend is in agreement with observations, but the amplitude of the signal depends strongly on how shapes are measured and how galaxies are selected in the simulation. Disc galaxies show a tendency to be oriented tangentially around spheroidals in three-dimensions. While this signal seems suppressed in projection, this does not guarantee that disc alignments can be safely ignored in future weak lensing surveys. The shape alignments of luminous galaxies in Horizon-AGN are in agreement with observations and other simulation works, but we find less alignment for lower luminosity populations. We also characterize the systematics of galaxy shapes in the simulation and show that they can be safely neglected when measuring the correlation of the density field and galaxy ellipticities.

Key words: cosmology: theory — gravitational lensing: weak – large-scale structure of Universe — methods: numerical

1 INTRODUCTION

There is mounting observational evidence that galaxies are subject to 'intrinsic alignments', i.e. correlations of their shapes across large separations due to tidal effects that act to align them in preferential directions with respect to one another (Brown et al. 2002; Aubert, Pichon & Colombi 2004; Mandelbaum et al. 2006a; Hirata et al. 2007; Joachimi et al. 2011; Heymans et al. 2013; Singh, Mandelbaum & More 2015). These alignments have been identified as an important systematic to weak lensing measurements, with the potential to undermine its capabilities as a probe of precision cosmology if unaccounted for (Hirata & Seljak 2004, 2010; Bridle & King 2007; Kirk et al. 2012; Krause, Eifler & Blazek 2015). To fully extract cosmological information from weak lensing, the alignment signal needs to be mitigated or marginalized over (Zhang 2010; Joachimi & Bridle 2010; Joachimi & Schneider 2010; Troxel & Ishak 2012). On the other hand, there is cosmological information to be extracted from the alignment signal (Chisari & Dvorkin 2013; Chisari, Dvorkin & Schmidt 2014; Schmidt, Chisari & Dvorkin 2015). For these reasons, this field is emerging as an interesting avenue for improving our understanding of galaxy formation and the evolution of the large-scale structure of the Universe. For a set of comprehensive reviews on the topic of intrinsic alignments, see Troxel & Ishak (2015), Joachimi et al. (2015), Kirk et al. (2015) and Kiessling et al. (2015).

The large-scale alignment signal of the shapes of luminous red galaxies (LRGs) with the matter density field has been considered the dominant contribution to low redshift intrinsic alignments and the main source of concern for weak lensing surveys. This signal has also been measured in numerical simulations of galaxy formation (Tenneti et al. 2014, 2015). A theoretical model developed by Catelan, Kamionkowski & Blandford (2001) suggested that elliptical galaxies, supported by random stellar motions, can be subject to large-scale tides, which stretch them along the direction of the tidal field. Alignments of LRGs

are well described by this model (Blazek, McQuinn & Seljak 2011).

Disc galaxies, which have significant angular momentum, can suffer a torque from the surrounding tidal field, which correlates their orientations (Catelan, Kamionkowski & Blandford 2001; Hui & Zhang 2002; Schaefer 2009; Schaefer & Merkel 2015). Hydrodynamical and N-body cosmological simulations have reported alignments of the spins of dark matter halos and galaxies with the cosmic web (Bailin & Steinmetz 2005; Aragón-Calvo et al. 2007; Hahn et al. 2007; Sousbie et al. 2008; Zhang et al. 2009; Hahn, Teyssier & Carollo 2010; Codis et al. 2012; Libeskind et al. 2012; Dubois et al. 2014) and there is indeed observational evidence of galaxy spin alignments (Pen, Lee & Seljak 2000; Paz, Stasyszyn & Padilla 2008; Jones, van de Weygaert & Aragón-Calvo 2010; Andrae & Jahnke 2011; Tempel & Libeskind 2013). Indirect evidence also comes from quasar polarizations, which have also been found to be aligned with the surrounding large-scale structure (Hutsemékers et al. 2014). However, at the same time, some works have claimed no alignment of discs (Slosar & White 2009; Hung & Ebeling 2012).

Disc galaxies have higher star formation than ellipticals, hence color information is typically used as a proxy to distinguish between types and alignment mechanisms. Despite evidence of their spin alignments, the shape alignment of blue (disc) galaxies has been observed to be consistent with null (Mandelbaum et al. 2011; Heymans et al. 2013). Recently, results from an analysis of the Horizon-AGN hydrodynamical simulation[1] (Codis et al. 2015) have suggested that this correlation of the shapes of blue galaxies could be present at high significance at redshift of $z = 1.2$, and it could thus become an additional contaminant to gravitational lensing. The main hypothesis behind that investigation was the use of spins as a proxy for galaxy shapes.

To elucidate the discrepancy between the non-detection in low-redshift observations of blue galaxy shape alignments and the prediction of significant spin alignments in simulations at high redshift, it is necessary to explore the relation between projected shapes and spins in the simulation in the redshift range where observational constraints are available. We therefore propose in this work to concentrate on the relationship between shapes and spins in the Horizon-AGN simulation (Dubois et al. 2014). In particular, we first explore the correlations between galaxy positions and orientations to gain insights into the physical processes that give rise to intrinsic alignments. We further measure projected correlations of galaxy positions and galaxy shapes, and auto-correlations of galaxy shapes, in a way that mimics the construction of intrinsic alignment correlations from observations. Note that all results are obtained at $z = 0.5$; the redshift evolution of this effect is left for future work.

In Section 2, we describe the hydrodynamical cosmological simulation used in this work to study the shape and spin alignments of galaxies, including the identification of galaxies in the simulation and the methods used to obtain their rest-frame colours. Section 3 discusses how shapes and spins are measured from the simulation and their convergence properties. In Section 4, we present the estimators used for correlation functions between spins, positions and shapes used in this work. We discuss both three-dimensional and two-dimensional (projected) correlations. While three-dimensional correlations give us an insight into the nature of the alignment process, the ultimate observables in photometric weak lensing surveys are the projected galaxy counts-ellipticity correlations. Section 5 brings together the results from the simulation and the theoretical modelling of the alignment signal. We discuss these results in relation to works in Section 6 and we conclude in Section 7.

Compared to previous work by Codis et al. (2015), in this work we lift the assumption that galaxy spins are a good proxy for galaxy shapes. Instead, we present results on alignments both for spins and shapes classifying galaxies by their kinematic properties. Moreover, we concentrate in correlations between positions and orientations, rather than correlations among galaxy orientations. This allows us to neglect potential systematics arising from spurious alignment correlations due to numerical issues (see Appendix A). Tenneti et al. (2015) measured projected shape correlations of galaxies in the smoothed-particle-hydrodynamics simulation MassiveBlack-II. This work presents shape and spin correlations measured in the Horizon-AGN simulation, which has a comparable volume to MassiveBlack-II but uses a different technique: adaptive-mesh-refinement. As the properties of galaxies might be sensitive to the numerical technique adopted, the comparison of alignments measured in Horizon-AGN and MassiveBlack-II is crucial to inform predictions of intrinsic alignment contamination to future weak gravitational lensing surveys.

2 THE SYNTHETIC UNIVERSE

In this section, we describe the Horizon-AGN simulation (see Dubois et al. 2014 for more details) and we explain how galaxy properties relevant to this study are extracted.

2.1 The Horizon-AGN simulation

The Horizon-AGN simulation is a cosmological hydrodynamical simulation in a box of $L = 100\,h^{-1}\,\mathrm{Mpc}$ a side. It is run adopting a standard ΛCDM cosmology compatible with the WMAP-7 cosmology (Komatsu et al. 2011), with total matter density $\Omega_\mathrm{m} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, amplitude of the matter power spectrum $\sigma_8 = 0.81$, baryon density $\Omega_\mathrm{b} = 0.045$, Hubble constant $H_0 = 70.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and $n_s = 0.967$. There are 1024^3 dark matter (DM) particles in the box, with a resulting DM mass resolution of $M_\mathrm{DM,res} = 8 \times 10^7\,\mathrm{M_\odot}$.

The adaptive-mesh-refinement (AMR) code ramses (Teyssier 2002) has been used to run the simulation with 7 levels of refinement up to $\Delta x = 1\,\mathrm{kpc}$. The refinement scheme follows a quasi-Lagrangian criterion. A new refinement level is triggered if the number of DM particles in a cell is more than 8 or if the total baryonic mass in a cell is 8 times the initial DM mass resolution.

Star formation is modelled following a Schmidt law: $\dot\rho_* = \epsilon_*\rho/t_\mathrm{ff}$, where $\dot\rho_*$ is the star formation rate density, ρ is the gas density, $\epsilon_* = 0.02$ (Kennicutt 1998; Krumholz & Tan

[1] `http://horizon-simulation.org`

2007) the constant star formation efficiency, and $t_{\rm ff}$ the local free-fall time of the gas. We allow star formation wherever the Hydrogen gas number density exceeds $n_0 = 0.1\,{\rm H\,cm^{-3}}$ according to a Poisson random process (Rasera & Teyssier 2006; Dubois & Teyssier 2008) with a stellar mass resolution of $M_{*,\rm res} = \rho_0 \Delta x^3 \simeq 2 \times 10^6\,{\rm M_\odot}$.

Gas cooling occurs by means of H and He cooling down to $10^4\,{\rm K}$ with a contribution from metals Sutherland & Dopita (1993). Following Haardt & Madau (1996), heating from a uniform UV background is implemented after the reionization redshift $z_{\rm reion} = 10$. Metallicity is modeled as a passive variable of the gas, which varies according to the injection of gas ejecta from supernovae explosions and stellar winds. We model stellar feedback using a Salpeter (1955) initial mass function with a low-mass (high-mass) cut-off of $0.1\,{\rm M_\odot}$ ($100\,{\rm M_\odot}$). In particular, the mechanical energy from supernovae type II and stellar winds follows the prescription of STARBURST99 (Leitherer et al. 1999, 2010), and the frequency of type Ia supernovae explosions is taken from Greggio & Renzini (1983).

Active Galactic Nuclei (AGN) feedback is modelled according to Dubois et al. (2012). We use a Bondi-Hoyle-Lyttleton accretion rate onto black holes, given by $\dot{M}_{\rm BH} = 4\pi\alpha G^2 M_{\rm BH}^2 \bar{\rho}/(\bar{c}_s^2 + \bar{u}^2)^{3/2}$, where $M_{\rm BH}$ is the black hole mass, $\bar{\rho}$ is the average gas density, $\bar{c}_s$ is the average sound speed, $\bar{u}$ is the average gas velocity relative to the black hole velocity, and α is a dimensionless boost factor. This is given by $\alpha = (\rho/\rho_0)^2$ when $\rho > \rho_0$ and $\alpha = 1$ otherwise (Booth & Schaye 2009) in order to account for our inability to capture the colder and higher density regions of the interstellar medium. The effective accretion rate onto black holes is not allowed to exceed the Eddington accretion rate: $\dot{M}_{\rm Edd} = 4\pi G M_{\rm BH} m_{\rm p}/(\epsilon_{\rm r}\sigma_{\rm T} c)$, where $\sigma_{\rm T}$ is the Thompson cross-section, c is the speed of light, $m_{\rm p}$ is the proton mass, and $\epsilon_{\rm r}$ is the radiative efficiency, assumed to be equal to $\epsilon_{\rm r} = 0.1$ for the Shakura & Sunyaev (1973) accretion onto a Schwarzschild black hole. Two different modes of AGN feedback are implemented: the *radio* mode, operating when $\chi = \dot{M}_{\rm BH}/\dot{M}_{\rm Edd} < 0.01$, and the *quasar* mode, active otherwise (see Dubois et al. 2014 for details).

2.2 Galaxy catalogue

Galaxies are identified in each redshift snapshot of HORIZON-AGN using the AdaptaHOP finder (Aubert, Pichon & Colombi 2004), which relies directly on the distribution of stellar particles. Twenty neighbours are used to compute the local density of each particle. To select relevant over-densities, we adopt a local threshold of $\rho_{\rm t} = 178$ times the average total matter density. Note that the galaxy population does not depend sensitively on the exact value chosen for this threshold. Our specific choice reflects the fact that the average density of galaxies located at the centre of galaxy clusters is comparable to that of the dark matter. The force softening is of approximately 10 kpc. Below this minimum size, substructures are treated as irrelevant.

Only galactic structures identified with more than 50 star particles are included in the mock catalogues. This enables a clear identification of galaxies, including those in the process of merging. At $z = 0.5$, there are $\sim 146\,000$ objects in the galaxy catalogue, with masses between 1.7×10^8 and $2.3 \times 10^{12}\,{\rm M_\odot}$. The galaxy stellar masses quoted in this paper should be understood as the sum over all star particles that belong to a galaxy structure identified by AdaptaHOP.

We compute the absolute AB magnitudes and rest-frame colours of the mock galaxies using single stellar population models from Bruzual & Charlot (2003) adopting a Salpeter initial mass function. Each stellar particle contributes a flux per frequency that depends on its mass, metallicity and age. The sum of the contribution of all star particles is passed through u, g, r, and i filter bands from the *Sloan Digital Sky Survey* (SDSS, Gunn et al. 2006). Fluxes are expressed as rest-frame quantities (i.e. that do not take into account the redshifting of spectra) and, for simplicity, dust extinction is neglected. Once all the star particles have been assigned a flux in each of the colour channels, we build the 2D projected maps for individual galaxies (satellites are excised with the galaxy finder). The sum of the contribution of their stars yields the galaxy luminosity in a given filter band.

3 SPINS AND SHAPES OF GALAXIES

To assign a spin to the galaxies, we compute the total angular momentum of the star particles which make up a given galactic structure relative to the centre of mass. We can therefore write the intrinsic angular momentum vector $\mathbf{L}$ or spin of a galaxy as

$$\mathbf{L} = \sum_{n=1}^{N} m^{(n)} \mathbf{x}^{(n)} \times \mathbf{v}^{(n)}\,, \tag{1}$$

where n denotes each stellar particle of mass $m^{(n)}$, position $\mathbf{x}^{(n)}$ and velocity $\mathbf{v}^{(n)}$ relative to the center of mass of that galaxy. The total stellar mass of a galaxy is given by $M_* = \sum_{n=1}^{N} m^{(n)}$. The specific angular momentum is $\mathbf{j} = \mathbf{L}/M_*$.

For each galaxy, we also obtain its V/σ, stellar rotation versus dispersion, that we measure from their 3D distribution of velocities. V/σ is a proxy for galaxy morphology: low values of this physical quantity indicate that a galaxy is more elliptical, pressure-supported by random stellar motions; high values of V/σ suggest a disc-like galaxy, with stellar rotation predominantly on a plane. We first compute the total angular momentum (spin) of stars in order to define a set of cylindrical spatial coordinates (r, θ, z), with the z-axis oriented along the galaxy spin. The velocity of each individual star particle is decomposed into cylindrical components v_r, v_θ, v_z, and the rotational velocity of a galaxy is $V = \bar{v}_\theta$, the mean of v_θ of individual stars. The velocity dispersion of each velocity component σ_r, σ_θ, σ_z is computed and used for the average velocity dispersion of the galaxy $\sigma^2 = (\sigma_r^2 + \sigma_\theta^2 + \sigma_z^2)/3$.

Figure 1 shows the distribution of V/σ for galaxies in the HORIZON-AGN simulation. We see an increase of the ratio V/σ from low-mass $M_* \simeq 10^9\,{\rm M_\odot}$ to intermediate-mass $M_* \simeq 2\times10^{10}\,{\rm M_\odot}$ galaxies, for which this median of the ratio peaks at 1. For massive galaxies, above $2\times10^{10}\,{\rm M_\odot}$, the V/σ ratio decreases with stellar mass. Dwarf and massive galaxies have low V/σ and hence are pressure-supported, while intermediate-mass galaxies have wide-spread V/σ values and are thus a hybrid population of rotation- and pressure-supported galaxies.

We also measure the inertia tensor of a galaxy to characterize its three dimensional shape. This tensor is given by

$$I_{ij} = \frac{1}{M_*} \sum_{n=1}^{N} m^{(n)} x_i^{(n)} x_j^{(n)} \,. \tag{2}$$

and it is then diagonalised to obtain the eigenvalues $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3$ and the corresponding unit eigenvectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ (respectively minor, intermediate and major axis of the ellipsoid). Analogously, we will explore the differences in measuring galaxy shapes using Equation (2) compared to the *reduced* inertia tensor, which is defined by

$$\tilde{I}_{ij} = \frac{1}{M_*} \sum_{n=1}^{N} m^{(n)} \frac{x_i^{(n)} x_j^{(n)}}{r_n^2}, \tag{3}$$

where r_n^2 is the three dimensional distance for the stellar particle n to the center of mass of the galaxy. The reduced inertia tensor is a closer representation of the shape of a galaxy as measured for weak gravitational lensing measurements, where the inner (and more luminous) region is upweighted with respect to the outskirts. For the reduced inertia tensor, Tenneti et al. (2014) found that the iterative procedure reduces the impact of the spherically symmetric r^{-2} weights, yielding shapes that are not as round as for the non-iterative procedure. We will consider here the simple and the reduced inertia tensor cases, and we expect that the results from applying iterative procedures to define the galaxy shapes (Schneider, Frenk & Cole 2012; Tenneti et al. 2014) will lie between the two cases considered.

Projected shapes are obtained by summing over $i, j = 1, 2$ in Equations (2) and (3), and the semiminor and semimajor axes are correspondingly defined by the eigenvectors of the projected inertia tensors. The axis ratio of the galaxy, $q = b/a$, is defined from the eigenvalues of the projected inertia tensor as the ratio of the minor to major axes ($b = \sqrt{\lambda_b}, a = \sqrt{\lambda_a}$, where λ_a is the largest eigenvalue and λ_b, the smallest). The components of the complex ellipticity, typically used in weak lensing measurements, are given by

$$(e_+, e_\times) = \frac{1-q^2}{1+q^2}[\cos(2\phi), \sin(2\phi)]\,, \tag{4}$$

where ϕ is the orientation angle, $+$ indicates the radial component of the ellipticity and $\times$ is the 45 deg-rotated component. Intrinsic alignments typically manifest themselves as a net average e_+ ellipticity around over-densities. The total ellipticity of a galaxy is thus $e = \sqrt{e_+^2 + e_\times^2}$. In this work, radial alignments have negative e_+ and tangential alignments (as expected from weak gravitational lensing) correspond to positive e_+. No net correlation of the $\times$ component with the matter density field is expected, and this statistic is thus used to test for systematics in the ellipticity measurement procedures. We discuss other systematics tests of ellipticity correlations in detail in Appendix A.

3.1 Convergence tests

Poorly resolved galaxies in the simulation are subject to uncertainties in their shapes and orientations. We identify two sources of uncertainties. First, there is the inherent variance in the shapes that arises from the choice of particles used in the shape computation. We call this measurement noise, $\sigma_{\rm meas}$, in analogy to the shape measurement from galaxy images for weak lensing. Second, there is a bias in the ellipticity and orientation measurement associated to the resolution of a galaxy, $\sigma_{\rm res}$. To define the minimum number of particles needed to obtain the shape of a galaxy, we compare these two uncertainties to the shape noise that arises from the dispersion in the intrinsic distribution of the shapes, σ_e.

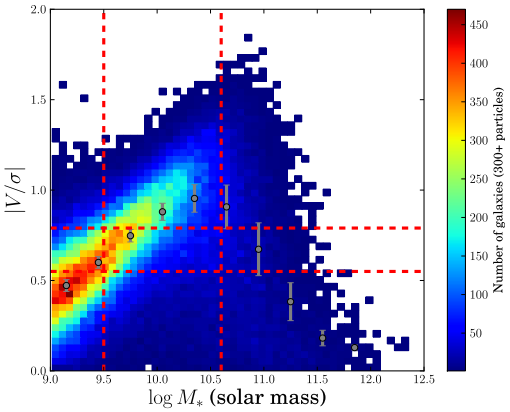


Figure 1. V/σ as a function of mass for the galaxies used in this work. The gray circles represent the median and the variance in 10 logarithmic bins of stellar mass. Dwarf and massive galaxies are pressure-supported, while intermediate-mass have a mixture of rotation and pressure support. The vertical red dashed lines represent the cuts corresponding to the different mass bins as in Codis et al. (2015). The horizontal red dashed lines represent our cuts in V/σ, chosen such that there is approximately the same number of galaxies in each V/σ bin.

The resolution bias is determined by randomly subsampling stellar particles in each galaxy with > 1000 particles. We compare the ellipticity measured from random subsamples of $50, 100, 300$ and 1000 stellar particles for those galaxies. The results are shown in Figure 2 and they suggest that a minimum number of 300 particles in each galaxy has to be required in order to guarantee that the bias in the ellipticity is an order of magnitude below the shape noise: $\sigma_{\rm res}^2 \lesssim 0.1\,\sigma_e^2$. The distributions of galaxy ellipticities are shown in Figure 3.

We determine $\sigma_{\rm meas}$ by bootstrap resampling (100 times) the stellar particles used for defining the inertia tensor from the overall population of stellar particles that make up each galaxy. We compute this uncertainty for galaxies with > 50, > 100, > 300 and > 1000 particles. Figure 4 shows the distribution of uncertainty in the shape measurement (compared to the rms ellipticity of the galaxy sample with 1000 particles) for each minimum number of particles considered. The solid lines correspond to the shape measured using the simple inertia tensor and the dashed lines, to those measured using the reduced inertia tensor. The uncertainties are typically below the 10^{-3} level and they are smaller in the case of the reduced inertia tensor. Resampling has a larger impact on the simple inertia tensor, since there are few particles in the outskirts of each galaxy, and these contribute equally to the shape measurement as those particles in the central part. Nevertheless, $\sigma_{\rm meas}$ is very small compared to the shape noise and to the resolution bias, and hence can be neglected. Notice that in our modelling we do

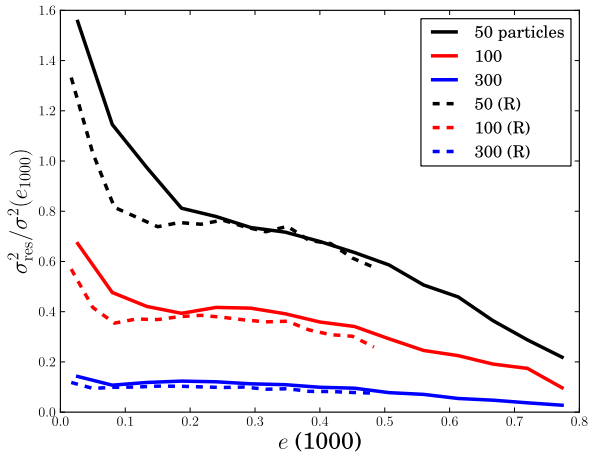


Figure 2. Resolution bias, $\sigma^2_{\rm res} = \langle|e_{50,100,300} - e_{1000}|^2\rangle/\sigma^2(e_{1000})$, as a function of the ellipticity obtained from the 1000 particle subsample for galaxies with more than 1000 particles. The colours indicate the number of stellar particles in the subsample: > 50 (black), > 100 (red) and > 300 (blue). The solid lines represent the results for the simple inertia tensor and the dashed lines, for the reduced inertia tensor. Subsampling has a larger impact on the simple inertia tensor due to the fact that particles in the outskirts contribute with equal weights as particles in the center.

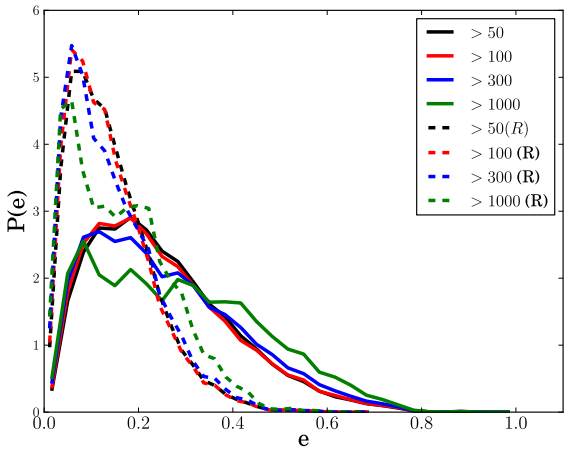


Figure 3. Distribution of ellipticities for galaxies with > 50 (black), > 100 (red), > 300 (blue) and > 1000 particles (green) for the simple (solid) and the reduced (dashed) inertia tensors. The shapes obtained with the reduced inertia tensor are rounder than those obtained with the simple inertia tensor.

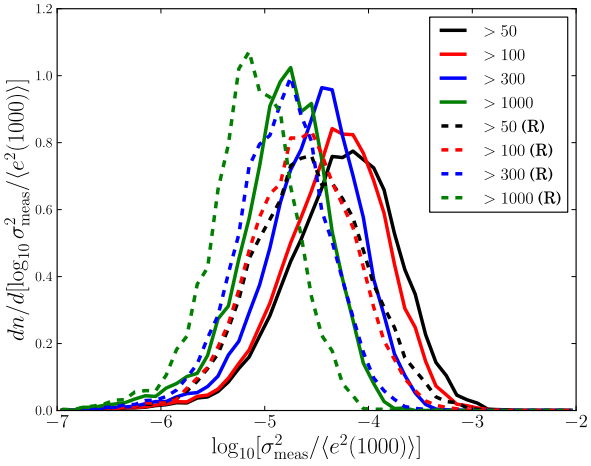


Figure 4. Distribution of measurement uncertainties in the ellipticities for galaxies with > 50 (black), > 100 (red), > 300 (blue) and > 1000 particles (green) for the simple (solid) and the reduced (dashed) inertia tensors. The shapes obtained with the reduced inertia tensor have significantly smaller uncertainties. These uncertainties are insignificant compared to the resolution bias and to the shape noise in the sample.

not include surface brightness cuts, attenuation by dust, the effect of the atmosphere or convolution by telescope optics. Also, the resampling of stellar particles is carried out over all stellar particles, which constitute a correlated data set. All of these could have a significant impact in the estimation of $\sigma_{\rm meas}$ but their modelling is outside the scope of this work.

We also estimated the impact of subsampling and resampling on the measurement of the specific angular momentum of the galaxies. We found that the measurement noise is insignificant and that the specific angular momentum is less sensitive to the number of particles in the subsample than the projected shapes. While subsampling introduces significant scatter in the measurement of the specific angular momentum, there is a strong correlation between this quantity measured with 50, 100, 300 and 1000 stellar particles. This correlation is shown for 50 versus 1000 particles in Figure 5.

We conclude that the minimum number of particles required to perform spin and shape measurements is determined by the requirements placed on the projected shapes. A minimum of 300 particles is needed for the uncertainty in projected shape to be at least one order of magnitude smaller than the shape noise. From this section onwards, we will work only with galaxies that have more than 300 stellar particles in the simulation. Notice that Tenneti et al. (2015) adopted a cut on 1000 stellar particles, while Velliscig et al. (2015) adopt similar cuts as in our work on the number of stellar particles. Figure 6 shows a comparison between the axis ratios of galaxies in the simulation using the simple (y-axis) and the reduced (x-axis) inertia tensors for galaxies with more than 300 stellar particles. The use of the reduced inertia tensor clearly results in rounder shapes as a result of up-weighting the inner regions of galaxies.

4 CORRELATION FUNCTIONS

4.1 Correlations in three dimensions

Three dimensional correlations of galaxy shapes and spins can give interesting insights into the formation processes leading to alignments. In this section, we define the correlation functions of galaxy *orientations* in three dimensional space. We define the orientation-separation correlation as a function of comoving separation r,

$$\eta_e(r) = \langle|\hat{\mathbf{r}}\cdot\hat{\mathbf{e}}(\mathbf{x}+\mathbf{r})|^2\rangle - 1/3\,, \quad (5)$$

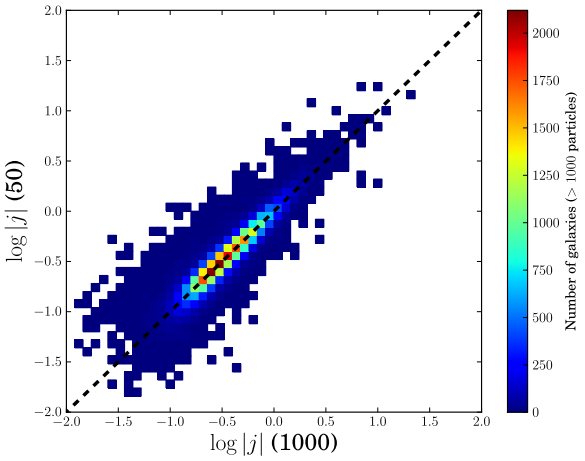


Figure 5. Correlation between the logarithm of the specific angular momentum, $j = |\mathbf{L}|/M_*$, when measured with 50 (y-axis) and 1000 (x-axis) particles, for galaxies in the simulation with > 1000 particles.

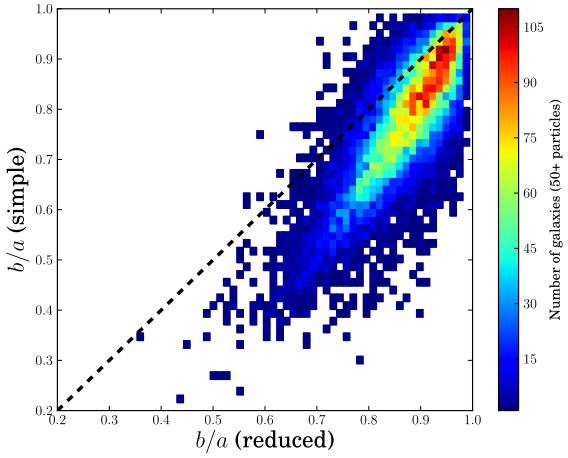


Figure 6. Axis ratios for galaxies with > 50 particles obtained with the simple inertia tensor (y-axis) an the reduced inertia tensor (x-axis). Colours indicate the number of galaxies in each bin. While there is a clear correlation between the estimates obtained with the two different methods, the reduced inertia tensor tends to produced rounder shapes.

where $\hat{\mathbf{e}}$ is the unit eigenvector of the inertia tensor pointing in the direction of the minor axis and $\hat{\mathbf{r}}$ is the unit separation vector. A positive correlation indicates a tendency for the separation vector and the minor axis of a galaxy to be parallel, hence for the galaxy to be elongated tangentially with respect to another galaxy. Notice that gravitational lensing produces a similar correlation for galaxy pairs separated by large distances along the line of sight. A negative correlation corresponds to a preferential perpendicular orientation of the minor axis with respect to the separation vector, resulting in a net radial orientation of galaxy shapes around other galaxies. Our treatment of galaxy orientations corresponds to representing the galaxy with the normal of its orientation plane and does not take into account the intrinsic dispersion in ellipticities. In other words, both spheroidals and discs are infinitely thin in this approach, and the plane that represents them points in the direction perpendicular to $\hat{\mathbf{e}}$. Alternatively, we will also measure the analogous correlation of separation vectors with spins, $\hat{\mathbf{s}}$, defined from the angular momentum of a galaxy, $\eta_s(r)$.

Grid-locking (studied in detail in Appendix A) manifests itself as a correlation between spin and shapes with the directions of the simulation box. As a result, spin and ellipticity auto-correlations (presented in Appendix B) can be contaminated by this effect, as two nearby galaxies formed from the same stream of infalling gas will be correlated among themselves and with the grid in which forces are evaluated and the effect can be enhanced at small distances because of the correlation length of the force field. On the contrary, we demonstrate that position-shape, position-spin and position-orientation correlations presented in the main body of this work are not affected.

4.2 Projected correlations

Three dimensional information of galaxy shapes and orientations is costly and near-term imaging surveys will be limited to correlations between projected shapes in broad tomographic bins. In this section, we describe projected statistics of alignments. First, we determine the orientation and ellipticity (Equation 4) of a galaxy from the projected inertia tensor. We obtain the real-space correlation function of galaxy positions and shapes projecting along one of the coordinates of the simulation box in a way that mimics the construction of Landy-Szalay estimators (Landy & Szalay 1993) in imaging surveys.

We label the dataset of tracers of the density field by D, the set of galaxies with ellipticities by S_+ (for the tangential ellipticity component, and analogously for the $\times$ component) and the set of random points by R. We define the correlation function of galaxy positions and ellipticities, $\xi_{g,+}(r_p, \Pi)$ as a function of projected separation in the sky, r_p, and along the line of sight, Π. This function is estimated from the sample of galaxies by using a modified Landy-Szalay estimator (Mandelbaum et al. 2011)

$$\xi_{g,+}(r_p, \Pi) = \frac{S_+D - S_+R}{RR}\,, \tag{6}$$

$$S_+D = \sum_{(r_p,\Pi)} \frac{e_{+,j}}{2\mathcal{R}}\,, \tag{7}$$

where $\mathcal{R}$ is the responsivity factor (Bernstein & Jarvis 2002), $\mathcal{R} = 1 - \langle e^2 \rangle$, $e_{+,j}$ is the tangential/radial component of the ellipticity vector of galaxy j and the sum is over galaxy pairs in given bins of projected radius and line of sight distance. S_+R is the sum of galaxy ellipticities around random points. In the previous expression, we are implicitly assuming that there is a single random sample, a procedure which is valid when the galaxy sample and the shape sample coincide. For the random sample, we use a set of points uniformly distributed in the simulation box with 10 times the density of the galaxy sample. For cross-correlations of different samples, generalization of Equation (6) is straightforward (e.g. Singh, Mandelbaum & More 2015).

This correlation function is then projected along the

line of sight by integration between $-\Pi_{\rm max} < \Pi < \Pi_{\rm max}$[2],

$$w_{g+}(r_p) = \int_{-\Pi_{\rm max}}^{\Pi_{\rm max}} {\rm d}\Pi\, \xi_{g+}(r_p, \Pi)\,, \quad (8)$$

where we take $\Pi_{\rm max}$ to be half the length of the simulation box. We are also interested in the correlation between the density field and the galaxy shapes, $w_{\delta+}(r_p)$, for which the galaxy positions are replaced with the positions of randomly sampled DM particles in the simulation box.

Two sources of uncertainties can be identified when measuring projected correlation functions of positions and shapes. The first one is shot noise coming from the tracers of the density and shear field. The second source of uncertainty is cosmic variance from the limited volume of the simulation box. Given that the box is $100\,h^{-1}\,$Mpc on each side, we limit the measurement of correlations to up to a quarter the length of the box size, L. However, this does not guarantee that we will be free from cosmic variance for the measured separations. The uncertainties in the projected correlation functions, including both shape noise and cosmic variance from modes within the box, can be obtained from jackknife resampling over the simulation box (Hirata et al. 2004; Mandelbaum et al. 2006b). We divide the simulation box in cubes of $L/3$ a side. In each jackknife iteration we remove the tracers corresponding to one of the cubes and compute a new estimate of the correlation. This allows us to estimate the variance in the projected correlations including the effect of cosmic variance. By contrast, the shape noise-only variance underestimates the jackknife variance by a factor of $1-4$ on small scales ($<1\,h^{-1}\,$Mpc) and up to approximately an order of magnitude on the largest scales probed in this work. In comparison, the jackknife accounts for covariance between Π bins and for the finite volume of the simulation box. Nevertheless, due to the limited size of the box, it is expected that the error bars could still be underestimated on large scales. We have also confirmed that using more jackknife regions with $L/4$ length a side does not alter the main conclusions of this manuscript.

5 RESULTS

5.1 Relative orientations in three dimensions

In this section, we characterize three dimensional correlations of galaxy positions and orientations at $z \simeq 0.5$ to shed light into the physical processes that can lead to alignments. Since three dimensional shapes are costly (see Huff et al. 2013 for a potential method), we study on-the-sky galaxy-ellipticity correlations in Section 5.2.

We use all galaxies in the simulation with > 300 stellar particles divided into three subpopulations by their V/σ, which is obtained as described in Section 3. This allows us to determine whether the alignment trends depend on the dynamical properties of the galaxies. Disc galaxies are represented by large values of V/σ, while spheroidals have low V/σ. As mentioned in Section 1, different alignment mechanisms are expected to act on these different populations (Catelan, Kamionkowski & Blandford 2001). Disc galaxies are expected to become aligned with the large-scale tidal field due to torques in their angular momentum, while ellipticals might suffer stretching or accretion along preferential directions determined by large-scale tides.

The three V/σ bins are constructed to host 1/3 of the galaxy population in each bin ($V/\sigma < 0.55$, $0.55 < V/\sigma < 0.79$ and $V/\sigma > 0.79$). The left panel of Figure 7 shows the three-dimensional correlation function of minor axis, from the simple inertia tensor, and separation vector. Error bars correspond to the standard error on the mean in each bin. We are only interested in a qualitative comparison in this section and we thus neglect cosmic variance except when explicitly mentioned. We observe that the minor axis orientation-position correlation is only significant in the low V/σ bin, suggesting that only galaxies with spheroidal dynamics are subject to shape alignments. The results imply that these galaxies are elongated pointing towards each other, in agreement with the qualitative behaviour expected from the tidal alignment model (Catelan, Kamionkowski & Blandford 2001). (Notice that, by construction, the fraction of spheroidal galaxies at this redshift is 1/3.) Galaxies with $V/\sigma > 0.55$ do not have correlated positions and shapes. Moreover, low V/σ have a lower degree of correlation between the direction of their minor axis and the direction of their spin, and this is a monotonic function of V/σ. Finally, we have split the low V/σ population of galaxies by the mean stellar mass; we find that both low ($\log M_* < 9.5$, $\sim 20,000$ galaxies) and high mass ($\log M_* > 9.5$, $\sim 8,000$ galaxies) galaxies are subject to alignments, and that high mass galaxies are more strongly aligned. Spin alignments are not very significant among these populations, although there seems to be a small trend for galaxies being oriented perpendicularly to the separation, i.e., tangentially around other galaxies. This also seen in the reduced inertia tensor of low and intermediate V/σ galaxies.

We also divide the galaxy population into three bins of mass: low mass ($10^9 < M_* < 10^{9.5}\,{\rm M}_\odot$, $\sim 30,000$ galaxies), intermediate mass ($10^{9.5} < M_* < 10^{10.6}\,{\rm M}_\odot$, $\sim 40,000$ galaxies) and high mass ($M_* > 10^{10.6}\,{\rm M}_\odot$, $\sim 7,000$ galaxies). The low mass limit is determined by the required threshold in the number of stellar particles. We first note that with this selection, both low and high mass galaxies tend to have low V/σ, while intermediate mass galaxies have V/σ values more consistent with those of a disc-like population, as shown in Figure 1. In Figure 8, we show the correlation between spin or minor axis with the separation vector for the three mass bins. The left and middle panels show the minor axis-separation correlation constructed using the simple (left) and reduced (middle) inertia tensor; and the right panel shows the spin-separation correlation.The left panel of Figure 8 shows that there is a significant correlation of the minor axis of high-mass galaxies aligned perpendicular to the separation vector towards other high mass galaxies. In the case of the reduced inertia tensor, the significance of this signal is decreased, consistently with the fact

[2] Notice that for simplicity we do not include the effect of peculiar velocities in transforming galaxy redshifts into distance along the line of sight, nor do we incorporate photometric redshift uncertainties. Both effects decrease the significance of intrinsic alignment correlations; however, in this work, we are interested in extracting as much information as possible from the simulation and we opt to avoid diluting the alignment signal by including these effects in the modelling.

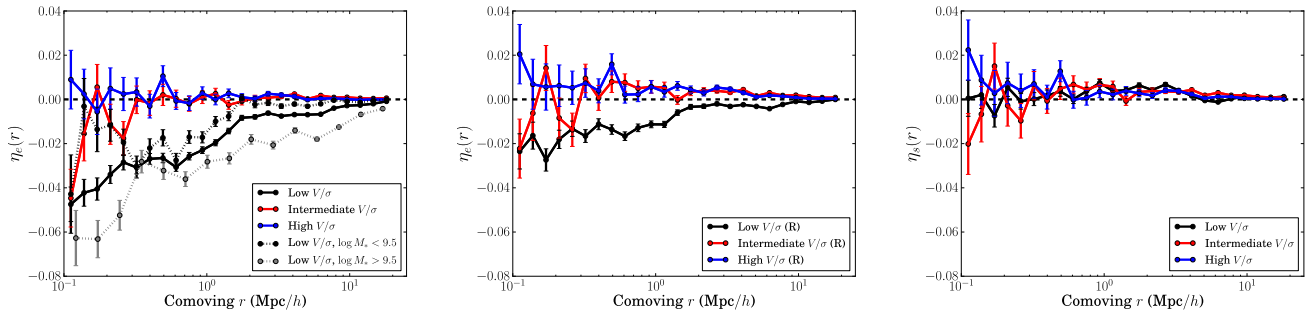


Figure 7. Correlations between the minor axis (η_e, left for simple inertia tensor and middle panel for reduced inertia tensor) or spin (η_s, right panel) and separation vector as a function of comoving separation. The sample of galaxies is divided into three bins by their dynamical classification: $V/\sigma < 0.55$, $0.55 < V/\sigma < 0.79$ and $V/\sigma > 0.79$. Galaxies with low V/σ show a significant shape-separation correlation. In the left panel, the dotted lines show the alignment signal for those galaxies split by the mean mass of that population: $\log M_* < 9.5$ (black) and $\log M_* > 9.5$ (gray). Both low and high mass populations have a contribution, and higher mass galaxies have stronger alignments.

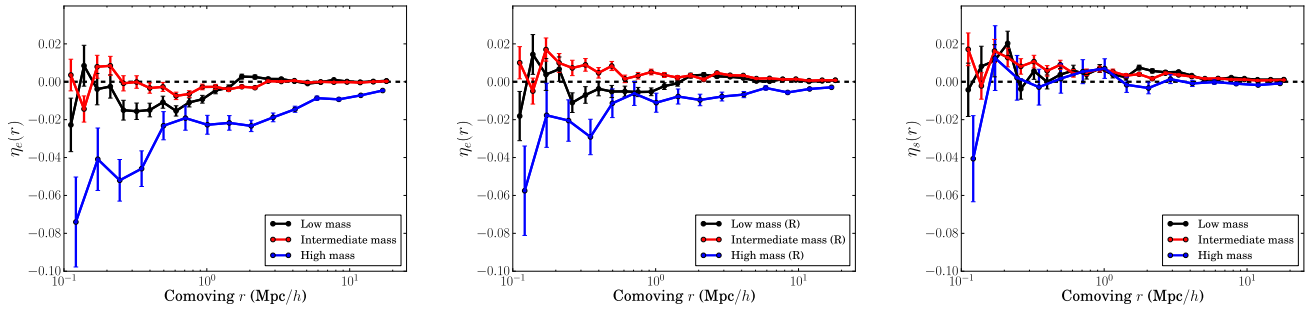


Figure 8. Correlation between the direction of the the minor axis (η_e) obtained from the simple inertia tensor (left panel) and from the reduced inertia tensor (middle panel) or spin of a galaxy (η_s, left panel) with the separation vector to a galaxy within the same mass-selected sample as a function of comoving separation. High mass galaxies show a clear radial alignment trend towards other high mass galaxies.

that the reduced inertia tensor yields rounder shapes for these galaxies. For low mass galaxies, the results are similar but with lower significance. For intermediate mass galaxies, we find a small negative correlation between the minor axis and the separation vector for the simple inertia tensor, and a small positive correlation when the reduced inertia tensor is used. This suggests that the reduced inertia tensor minor axis is correlated with the direction of the spin for this disc-like population, while this is not the case for the simple inertia tensor. We interpret this as a decreasing tendency of stellar particles to settle on a disc as a function of distance to the center of mass of the galaxy, possibly tracing merging structures. Indeed, satellite mergers within the Horizon-AGN simulation tend to redistribute their angular momentum significantly within the host (Welker et al., in prep.).

The overall physical picture that we get from these results is the following.

- Spin and shape alignments depend on galaxy dynamics, and these trends also translate into a mass dependence.
- Spheroidal galaxies show a significant trend of radial alignments with respect to each other that is preserved to large separations and is more prominent for high mass galaxies. The signal decreases in amplitude when the reduced inertia shapes are adopted, as the galaxy shapes become rounder and less sensitive to tidal debris in the outskirts.
- Spin alignment trends are tangential around other galaxies and marginal, and seem to be better correlated with the reduced inertia tensor than with the simple inertia tensor shapes for intermediate and high V/σ galaxies (correspondingly, also intermediate mass galaxies).

The results presented so far only consider correlations between galaxies with similar properties. We also explore whether cross-correlations between subsets exist. Figure 9 shows the cross-correlation of positions of spheroidal tracers ($V/\sigma < 0.55$) with the direction of the minor axis (red circles) and the spins (red triangles) of disc-like tracers ($V/\sigma > 0.55$). Discs show a tendency to align their spins parallel to the separation vector to spheroidal galaxies (red dashed curve), as well as their minor axis (red solid curve). Notice that the direction of the spin and the minor axis of a disc galaxy (from the reduced inertia tensor) are very well correlated, and hence the red dashed and solid curves almost lie almost on top of each other. Figure 9 also shows the cross-correlation of the positions of discs with the orientations of the minor axis (black circles) and the spin (black triangles) of spheroidal galaxies. These galaxies show a preferential elongation of their shapes towards the positions of discs, while they do not show significant alignment of

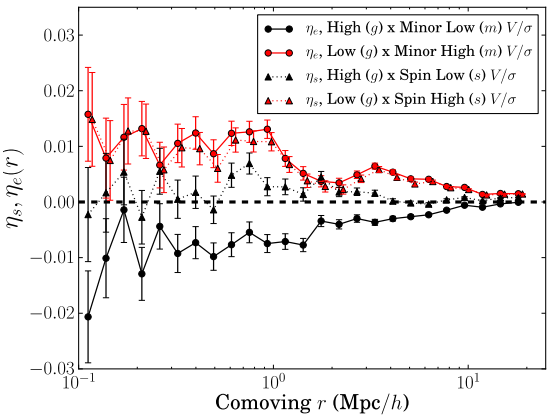


Figure 9. Cross-correlations of galaxy positions and shapes (η_e) or spins (η_s) for the galaxy population divided in two V/σ bins. δ indicates tracers of the galaxy density field and s (m) indicates the sample for which the orientation is measured from the spin (minor axis). The black solid line shows the correlation between the positions of disc-like tracers of the density field ($V/\sigma > 0.55$) and the direction of the minor axis of the reduced inertia tensor of spheroidal galaxies ($V/\sigma < 0.55$). The red solid line shows the correlation between the positions of spheroidal galaxies and the direction of the minor axis of disc-like galaxies. The red dotted line shows the correlation between the position of discs and the spin of spheroidals, and the red line shows the correlation between the positions of spheroidals and the direction of the spin axis of discs. (Notice that the red lines almost overlap, showing that the spin and the minor axis of a disc point along the same direction.) These results suggest that the discs are preferentially clustered in the direction of the elongation of spheroidals, while they also tend to have their spins aligned in the direction of nearby spheroidals.

their spins. When the simple inertia tensor shapes are used, the spins and shapes of discs become decorrelated and the significance of the shape alignment trend of discs around spheroidals is lost. From these results, we conclude that the discs are preferentially clustered in the direction of the elongation of spheroidals, while they also tend to have their spins aligned in the direction of nearby spheroidals (also traced by their reduced inertia tensor shapes). This picture is in agreement with disc galaxies living predominantly in a filamentary structure that follows the elongation of spheroidal galaxies at its knots. Furthermore, disc galaxies tend to have their spins aligned parallel to the direction of these filaments, and perpendicular to the elongation of the central spheroidal. Figure 10 shows a cartoon picture of alignments where the effect is exaggerated for visual purposes.

We perform a series of tests to determine the significance of the measured signal of disc alignments with respect to spheroidal tracers. In particular, in this case we adopt a jackknife procedure (similar to that adopted for projected correlations) to estimate the uncertainty in the disc alignment signal. The significance of the measurement is obtained from a χ^2 test using only the diagonals of the jackknife covariance matrix and without modelling the noise in this matrix. We find that the null hypothesis can be rejected at $> 99.99\%$ confidence level (C.L.) level for the alignment of disc spins around spheroidals (red dashed in Figure 9), and similarly for the alignment of the minor axes of discs around spheroidals (red solid line in Figure 9). This level of significance decreases to 92% when the simple inertia tensor is adopted to measure the shapes. On the contrary, the direction of the spin of spheroidals is not correlated with the position of discs (72% C.L. for null hypothesis rejection of the blue curves), but the minor axis direction of a spheroidal is anti-correlated with the position of discs with high significance ($> 99.99\%$ C.L. for both the reduced ans simple inertia tensor). Finally, we consider whether alignment signals are still present when the orientation is defined by the direction of the *major* axis. We find that the disc alignment measurement in this case is more sensitive to the choice of reduced/simple inertia tensor. This result confirms that the simple inertia tensor is a worse tracer of the spin compared to the reduced inertia tensor, as the alignment signal loses significance in that case. On the other hand, the use of the simple or reduced inertia tensor does not change the significance of the alignment of spheroidals in the direction of discs.

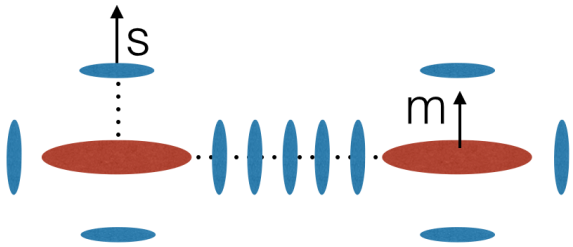


Figure 10. A cartoon picture of alignments, as interpreted from the results of Section 5.1. Discs live in filaments connecting ellipticals and they tend to align their spin(**s**)/minor axes in the direction of the filament. Ellipticals tend to have their shapes (**m** represents the minor axis) aligned towards each other and towards the direction of the filaments. The effect of alignments is exaggerated for visual purposes by showing all galaxies perfectly aligned following the measured trends in the simulation.

We conclude that:

- discs show a significant tendency for tangential alignment around over-densities traced by spheroidal galaxies,
- spheroidals are preferentially elongated towards discs and other spheroidals,
- and that spin is a good tracer of reduced inertia shapes for discs, but not for spheroidals.

In the next section, we mimic observations by exploring projected ellipticity alignments.

5.2 Projected correlations

The intrinsic alignment signal is typically measured in the literature using the projected correlation function of galaxy positions and shapes (Equation 8). This quantity is readily accessible using shear measurements from survey galaxy catalogs. In this work, we also obtain the projected correlation functions of the density field and projected shapes, $w_{\delta+}$ and $w_{\delta\times}$, where the density field is obtained from a random subsampling of 0.007% of the DM particles in the box. This subsampling guarantees a 10% convergence level in the DM power spectrum, which is similar to the expected level of convergence in determining galaxy shape (as we discussed

from Figure 2). Tenneti et al. (2015) used a similar approach with a comparable subsampling fraction. The measurement of $w_{\delta+}$ and $w_{\delta\times}$ is advantageous in that it allows us to avoid modelling galaxy bias, or to make any assumption about its scale dependence. Also, we do not need to model peculiar velocities, as the DM and galaxy positions in the box are perfectly known.

We measure the $w_{\delta+}$ correlation function following Equation (8) for all galaxies with > 300 stellar particles in the simulation box and replacing the density tracers by the subsample of DM particles. Grid locking (see Appendix A) is not expected to contaminate the measurement by spherical symmetry. As a consequence, the S_+R term is not expected to contribute to this correlation and we neglect it in this section. Section A2 provides confirmation of these assumptions. We show the projected correlation functions of the density field and the + component of the shape from the simple inertia tensor and the reduced inertia tensor in the left panel of Figure 11. As expected from our results in Section 5.1, we find an anti-correlation between the + component of the shape and the density field that is significant at $> 99.99\%$ C.L. level for the simple inertia tensor. The negative sign indicates that the projected shapes of galaxies are elongated pointing towards other galaxies, i.e., that alignments are radial. We find a decreased tendency for alignments (47% C.L.) when using the reduced inertia tensor, consistently with rounder shapes and with the results presented in Section 5.1. The right panel of Figure 11 shows that the $\delta\times$ correlation is consistent with null (at the $\simeq 65\%$ C.L.).

In Figure 12, we split the sample of galaxies with shapes into 5 bins of mass (left panel), V/σ (middle panel) and $u-r$ colour (right panel). All bins have approximately the same number of galaxies and the legend in each panel indicates the mean of the property considered for the galaxies in each bin. We find that galaxies in the lowest (highest) V/σ ($u-r$ colour) bin carry the strongest alignment signal. We find very similar results when splitting the galaxies by their $g-r$ colour. In comparison, the split by mass results in a less clear identification of which galaxies are responsible for projected shape alignments. Figure 13 shows that there is a significant correlation between colour and V/σ for galaxies with redder colours. On the contrary, V/σ and mass are not monotonically correlated, as shown earlier in Figure 1.

Interestingly, the high V/σ galaxies do not show any significant alignment in Figure 12. This is puzzling given the results presented in Figure 9. To elucidate this discrepancy, we compute the w_{g+} statistic using the same selection cuts as for Figure 9 and show the results in Figure 14. We find that, while the alignment of spheroids in the direction of the clustering of discs is still significant in projection, the disc alignment signal is diluted and consistent with null at the 27% C.L. using the simple inertia tensor, but less so (89% C.L.) using the reduced inertia tensor. In the latter case, the signal was more significant from the orientation-separation correlation of Section 5.1. In projection, a *tangential alignment* is only marginally present. (We remind the reader that gravitational lensing has the same positive sign for correlations measured between pairs of galaxies with large separations along the line of sight). Notice that, as discussed in Section 5.1, a lower level of correlation is expected for the simple inertia tensor given that this is not a good tracer of

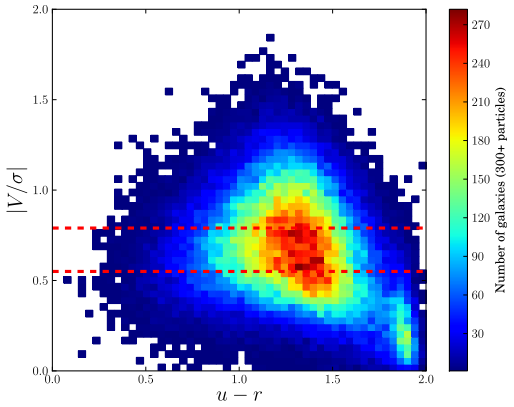


Figure 13. Galaxies in the V/σ vs $u-r$ plane. There is a strong correlation between V/σ and $u-r$ colour at the red end of the colour distribution. Galaxies in Horizon-AGN have a bimodal distribution of colours and V/σ. The red horizontal line represent our fiducial cuts in V/σ used in Section 5.1.

spin alignment. Moreover, there are several reasons why the disc orientation correlation observed in Figure 9 can be diluted in projection. One factor is the weighting by galaxy ellipticity in Equation (7). Face-on discs would carry no signal in this statistic. The second reason for dilution is the fact that Equation (8) weights the signal in each Π bin equally, while the alignment signal is expected to lose correlation as Π increases. In comparison, Figure 9 showed the level of alignment as a function of three dimensional separation: pairs with large Π have large r in that figure and lower correlation.

Finally, we also study the cross correlation of galaxy positions and galaxy shapes, considering all galaxies in the simulation box. This correlation will include the effect of galaxy bias, compared to the DM-shape correlation. We show a comparison of cross correlation of + shapes and DM, and of + shapes and galaxy positions in the left panel of Figure 11. We find that the galaxy-shape correlation traces the DM-shape correlation well within the error bars. There is a discrepancy in amplitude of the correlation in the first bin in the case of the simple inertia tensor. The excess power in the galaxy-shape correlation could arise from tidal debris on small scales or from the increased clustering compared to the DM, but we cannot draw firm conclusions from this comparison. In general, the similarity between the galaxy-shape correlation and the DM density-shape correlation suggests that the bias parameter b_g is not very different from unity for the sample of galaxies considered, and that it does not have any significant scale dependence.

5.3 Modelling of alignment signal

Early-type galaxy alignments are thought to arise due to the action of the tidal field. In this model, the tidal field contributes a small component to the projected ellipticity of a galaxy, given by (Catelan, Kamionkowski & Blandford 2001)

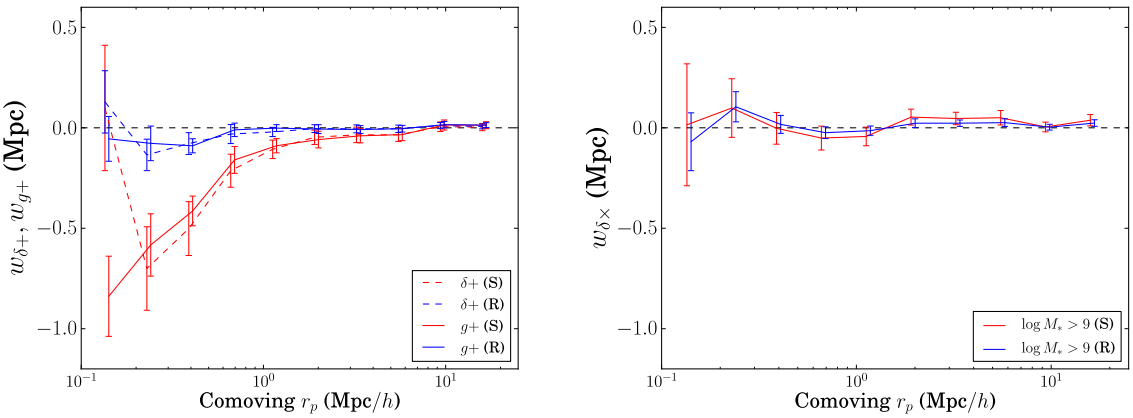


Figure 11. $w_{\delta+}$ projected correlation function for all galaxies with > 300 stellar particles (left panel). This cut corresponds to a cut in mass of $\log(M_*/\mathrm{M}_\odot) > 9$. The same panel also presents a comparison between the $w_{\delta+}$ correlation and the correlation between galaxy positions and + component of the shapes w_{g+}. The right panel shows the projected $\delta\times$ correlation as a test for systematics. In both panels, results obtained with the simple inertia tensor are indicated with the blue line; while the red line corresponds to shapes obtained from the reduced inertia tensor. The measured points for w_{g+} are arbitrarily displaced to larger radii by 5% for visual clarity in the left panel, and similarly for the reduced inertia tensor in the right panel.

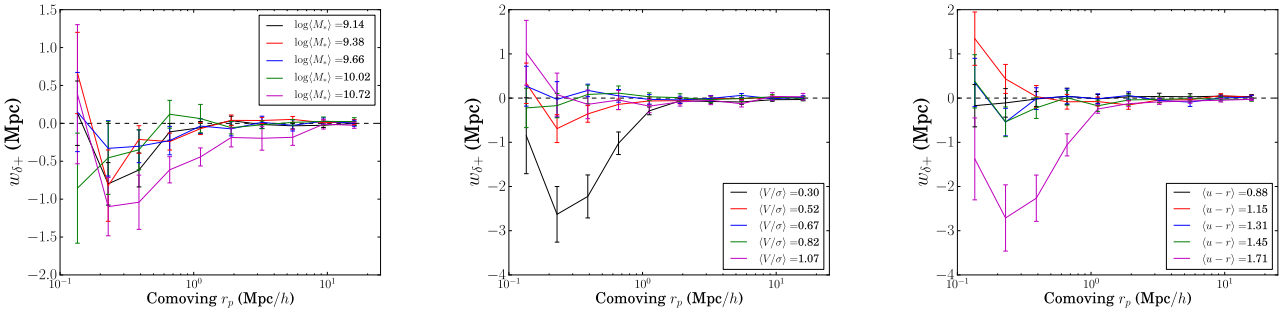


Figure 12. $w_{\delta+}$ projected correlation functions for 5 bins of mass (left panel) and V/σ (middle panel) and $u-r$ colour (right panel). The legend indicates the mean of the considered property in each bin, and the 5 bins are approximately equally populated. For simplicity, we only show the correlations that correspond to shape measurements using the simple inertia tensor. The impact of using the reduced inertia tensor is shown in Figure 11.

$$\gamma^I_{(+,\times)} = \frac{C_1}{4\pi G}(\partial_x^2 - \partial_y^2, \partial_x\partial_y)\mathcal{S}[\phi_p], \tag{9}$$

where C_1 is a proportionality constant that parametrizes the response of the shape of a galaxy to the tidal field, ϕ_p is the Newtonian gravitational potential at the redshift of formation of a galaxy and $\mathcal{S}$ is a smoothing filter that acts to smooth the potential over the typical scale of the galactic halo (~ 1 Mpc)[3]. As a consequence, there is a correlation between galaxy positions and their intrinsic shapes, given by

$$P_{g+}(\mathbf{k}, z) = -\frac{b_g C_1 \rho_{\rm crit}\Omega_m}{D(z)}\frac{k_x^2 - k_y^2}{k^2}P_\delta(\mathbf{k}, z)\,, \tag{10}$$

where b_g is the galaxy bias, $\rho_{\rm crit}$ is the critical density of the Universe today, $D(z)$ is the growth function (normalized to unity at $z = 0$) and P_δ is the matter power spectrum. While these expressions are not strictly valid in the non-linear regime, it is customary to approximate the nonlinear scale alignments by replacing the linear power spectrum in Equation (10) (Bridle & King 2007) by its nonlinear analogue. Recent observational works are beginning to test this assumption (Singh, Mandelbaum & More 2015), but most constraints on the amplitude of alignments are still typically given in terms of the nonlinear alignment ('NLA') approximation of Bridle & King (2007). The $g\times$ power spectrum is not presented because it is expected to average to null in projection.

[3] Notice that we adopt a different sign convention than Singh, Mandelbaum & More (2015) and Tenneti et al. (2015), whereby alignments are negative if they are radial, and positive if tangential.

Equation (10) can be transformed to redshift space to give a prediction for the on-the-sky w_{g+} projected correlation,

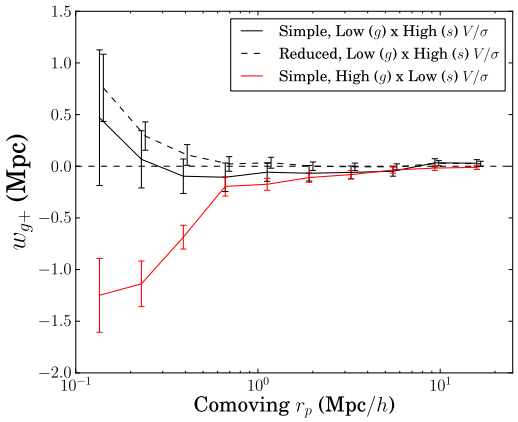


Figure 14. Cross-correlation of galaxy positions and + component of the shape applying the same selection as in Figure 9. The alignment of low V/σ galaxies in the direction of high V/σ galaxies (red) is significant after projection of the shapes and through the simulation box. On the contrary, the tangential alignment of high V/σ galaxies around low V/σ tracers of the density field (black) is diluted in projection, albeit the dashed line still rejects the null hypothesis at 89% C.L. The error bars in the case of the black dashed line have been artificially displaced to 5% larger r_p for visual clarity.

$$w_{g+}(r_p) = -\frac{b_g C_1 \rho_{\rm crit} \Omega_m}{\pi^2 D(z)} \int_0^\infty dk_z \int_0^\infty dk_\perp \frac{k_\perp^3}{(k_\perp^2 + k_z^2) k_z} P_\delta(\mathbf{k}, z) \sin(k_z \pi_{\rm max}) J_2(k_\perp r_p), \quad (11)$$

where b_g is the galaxy bias, r_p is the projected radius, $\Pi_{\rm max}$ is the line of sight distance over which the projection is carried out and J_2 is the second order Bessel function of the first kind. Notice that the Kaiser factor (Kaiser 1987; Singh, Mandelbaum & More 2015) is unnecessary to model the galaxies in the simulation, since we have access to the true positions of the galaxies along the line of sight. Similarly, we can safely neglect fingers-of-God effects arising from peculiar velocities in the nonlinear regime.

Singh, Mandelbaum & More (2015) observed an excess of power on small scale alignments compared to the best fit NLA model to the LOWZ galaxy sample of the SDSS. In this regime, we also consider a halo model developed by Schneider & Bridle (2010). According to this model, the one halo power spectrum of galaxy positions and intrinsic shapes is given by

$$P^{1h}_{\delta,\gamma^I}(k, z) = -a_h \frac{(k/p_1)^2}{1 + (k/p_2)^{p_3}}, \quad (12)$$

where a_h is the halo model alignment amplitude on small scales; p_1, p_2, and p_3 are fixed parameters based on fits by Schneider & Bridle (2010); and the projected correlation function is then given by

$$w^{1h}_{g+} = -b_g \int \frac{dk_\perp}{2\pi} k_\perp P^{1h}_{\delta,\gamma^I}(k_\perp, z) J_0(k_\perp r_p). \quad (13)$$

While the alignments of disc galaxy shapes have not been detected in observations, and despite the fact that the tidal alignment model is not expected to describe their alignments (Catelan, Kamionkowski & Blandford 2001; Hirata & Seljak 2004), the available constraints are typically phrased in terms of the NLA model as well (e.g., Mandelbaum et al. 2011).

We now place quantitative constraints on the goodness-of-fit of the intrinsic alignment models. We build a model template of the intrinsic alignment signal, $w^\mu_{\delta+}(r_p)$, at the positions for which we measure this correlation in the simulation.

Using the diagonals of the jackknife covariance, we determine the χ^2 from summing over all radial bins as follows

$$\chi^2 = \sum_{r_p} \frac{(w_{\delta+}(r_p) - w^\mu_{\delta+}(r_p))^2}{{\rm Var}[w_{\delta+}(r_p)]}, \quad (14)$$

where μ represents the alignment model template to be fit and we look for the minimum χ^2 by varying the parameters of the fit; and analogously for w_{g+}, in which case only joint constraints of the product of b_g and the alignment amplitude are obtained. We emphasize that we are not modelling covariances or noise in this matrix when performing the fits.

Notice that while the results presented in the previous section do not guarantee that disc-like galaxies do not contribute to the alignment signal, the correlation functions shown in Section 5.1 suggest that they counter-act the radial alignment of ellipticals due to their tendency to orient tangentially around over-densities. For this reason, and given that we do not have sufficient constraining power to bin the galaxy sample into different populations, we fit different models to w_{g+} and $w_{\delta+}$ shown in Figure 11, which include the contribution of all galaxies with $> 10^9 {\rm M}_\odot$ stellar masses in the simulation.

We focus on fits to the signal obtained using the simple inertia tensor, as in the case of the reduced inertia tensor the results are consistent with null. A power-law fit as a function of projected radius ($w^\mu_{g+} = b_g A_I r_p^\beta$) yields the following constraints on power-law amplitude and power-law index: $b_g A_I = -0.13 \pm 0.03$ and $\beta = -0.75 \pm 0.25$, respectively. The error bars quoted correspond to 68% C.L when holding one of the parameters fixed at the best fit value corresponding to the minimum χ^2 ($b_g A_I = -0.17$ and $\beta = -1.0$). We find that the linear alignment model is a very poor fit due to the lack of power on small scales compared to the measured signal. Similarly, the NLA model also underestimates the correlation of positions and shapes on small scales[4]. In this case, the best fit model is shown in Figure 15. The best fit NLA model for w_{g+} in the case of the simple inertia tensor is represented by the black solid line. In gray, we show the best fit NLA model for $w_{\delta+}$, also for the simple inertia tensor, which is comparable to the black line. The black dotted line shows the best fit to $w_{\delta+}$ in the case of the reduced inertia tensor, for which the best fit amplitude is consistent with null at 68% C.L.

Given that the best fit NLA model underestimates the alignment signal on small scales, we consider fitting a sum of the NLA model template (across all scales measured) and the halo model (only at $r_p < 0.8/h$) with the parameters of

[4] In this particular work, we do not apply any smoothing to the tidal field. A smoothing filter would suppress power on small scales, worsening the comparison of the NLA model to the measured signal.

Schneider & Bridle (2010). Notice that the physical interpretation of these results is not straightforward and should only be considered as phenomenological. The constraints on the parameters for the NLA and halo model combination are: $b_g a_h = 0.27^{+0.12}_{-0.11}$, $b_g A_I = 0.58^{+0.34}_{-0.35}$. The χ^2 in this case is comparable to that of the power-law fit. If we use more and smaller ($L/4$ a side) jackknife regions, the error bars increase slightly yielding $b_g A_I = 0.60^{+0.51}_{-0.41}$.

Given the limited simulation volume and the corresponding (cosmic) large error bars, we cannot obtain meaningful constraints on mass or luminosity dependence of the signal. However, we note that, as discussed in Section 5.2, the mass dependence is not monotonic.

6 DISCUSSION

Theoretical models of intrinsic alignments have indeed suggested that the population of galaxies subject to this mechanism can be split in two (Catelan, Kamionkowski & Blandford 2001). Disc-like systems interact with the large-scale structure through torques to their angular momentum vector, while spheroidals, which do not have significant angular momentum, would tend to orient their major axes pointing towards over-densities. In this work, we have used V/σ as a proxy of galaxy morphology, and we have indeed confirmed the existence of two different alignment mechanisms in play for spheroidals (low V/σ) and disc-like (high V/σ) galaxies. While V/σ is not easily accessible to upcoming gravitational lensing surveys, fortunately there is a strong correlation between $u-r$ and V/σ (Figure 13) that can help identify the two populations of galaxies and their alignments. Low V/σ galaxies tend to have redder colours, while high V/σ galaxies tend to be bluer.

The alignments of blue galaxy shapes have so far been consistent with null from observations (Mandelbaum et al. 2011; Heymans et al. 2013), albeit with large error bars that still allow for a significant level of contamination from blue galaxy alignments to cosmological observables in current and future surveys (Chisari et al. 2015; Krause, Eifler & Blazek 2015). We similarly find that the projected correlation function of blue galaxy positions and shapes is consistent with null in Horizon-AGN. Nevertheless, while blue galaxies do not align around each other, they tend to align tangentially around red galaxies in three dimensional space, and this signal is washed out in projection *within our error bars*. We emphasize that this does not imply that blue galaxy alignments can be neglected for future surveys. Codis et al. (2015) found that spin-spin alignments could potentially translate into worrisome levels of contamination for future lensing surveys, particularly for blue galaxies. Their results were based on a "spin-gives-ellipticity" prescription for translating spins into shapes for disc-like galaxies. We found that the validity of this assumption depends on the method used to determine ellipticity, with better agreement for the reduced inertia tensor case. The statistics of spin/shape alignments of disc galaxies in three dimensions from the simulation can be a useful tool to constrain disc alignment models and determine the level of contamination to future surveys.

Several theoretical predictions have been made for the spin alignments of disc-like galaxies. In simulations, discs tend to form with their spin correlated with the direction of filaments (Bailin & Steinmetz 2005; Aragón-Calvo et al. 2007; Hahn et al. 2007; Sousbie et al. 2008; Zhang et al. 2009; Codis et al. 2012; Libeskind et al. 2012; Dubois et al. 2014) and the vorticity of the density field around them (Laigle et al. 2015). This is consistent with the scenario of spin acquisition by tidal torquing biased by the large scale structure filaments. Extended tidal torque theory predicts that halos with mass below $5 \times 10^{12}\,\mathrm{M}_\odot$ at redshift $z = 0$ tend to form with their spin pointing along the direction of filaments, the center of which are represented by saddle points of the density field (Codis, Pichon & Pogosyan 2015). In the plane perpendicular to the filament and containing the saddle point, halos are predicted to show this preferential orientation. Away from this plane and closer to the nodes of the filament where higher mass halos reside, spins flip direction, becoming perpendicular to the filamentary axis. Hence, if galactic spins are correlated with halo spins, we should expect low-mass galaxies to have a spin aligned with the filamentary axis (as found in Dubois et al. (2014) for Horizon-AGN galaxies) and therefore also aligned[5] with the separation vector (since we expect a higher number of galaxies to inhabit within the filament). In contrast, Codis, Pichon & Pogosyan (2015) predict that very massive galaxies should have a spin perpendicular to the filament. However, massive galaxies tend to be supported by random motions of their stars, and hence the definition of the spin direction becomes more noisy. The signal predicted by theory could also be diluted due to the fact that not all low mass galaxies reside in filaments. We found that the spin and the minor axes of massive galaxies have less tendency to be aligned than for their lower mass counterparts.

Works by Tenneti et al. (2014, 2015); Velliscig et al. (2015) have also studied the intrinsic alignments of galaxies in hydrodynamical simulations. Tenneti et al. (2015) find that orientation-position correlations (their 'ED' correlation) with the density field have similar strengths and signs for blue and red galaxies when correlated with the density field. On the contrary, we find that blue galaxies tend to have *tangential alignments* around the locations of red galaxies (but not around each other). In projection, both works find the blue galaxy alignments to be suppressed. Those authors did not study spin alignments, which in this work have been shown to be significant for disc-like galaxies and to be connected to their shape alignments.

Observational results have shown there is a strong trend for radial alignment of the ellipticities of red galaxies (Mandelbaum et al. 2006a; Hirata et al. 2007; Okumura, Jing & Li 2009; Joachimi et al. 2011; Heymans et al. 2013; Singh, Mandelbaum & More 2015), which increases for higher mass galaxies. This trend is also seen in Horizon-AGN for the shapes of spheroidals. Tenneti et al. (2015) presented NLA and power-law fits to $w_{\delta+}$ from their simulation as a function of redshift and luminosity of the galaxy sample using iterative reduced inertia tensor for the galaxy shapes. We find that their fits reproduce the alignment signal of luminous galaxies

[5] or anti aligned, should the galaxies belong to octants of opposite polarity, see Codis, Pichon & Pogosyan (2015), section 6.2.3.

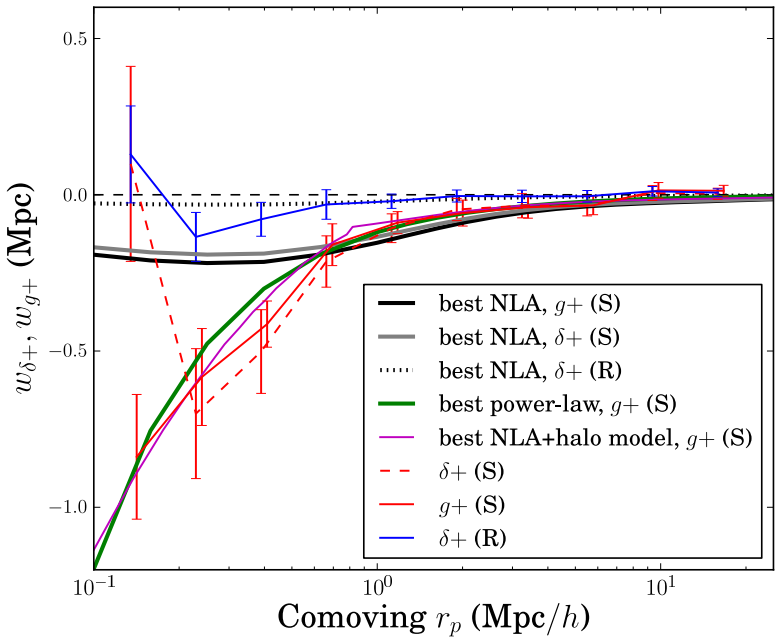


Figure 15. Alignment model fits for $w_{\delta+}$ and w_{g+} for all galaxies with > 300 stellar particles. Results obtained with the simple inertia tensor are indicated with the red lines; while the blue line corresponds to shapes obtained from the reduced inertia tensor for $w_{\delta+}$ only. We show fits from the NLA model to $w_{\delta+}$ (gray solid for simple inertia tensor; black dotted for reduced inertia tensor) and to w_{g+} (black solid for the simple inertia tensor). The measured points for $w_{\delta+}$ and simple inertia tensor shapes are arbitrarily displaced to larger radii by 5% for visual clarity. The NLA model significantly underestimates the power in alignments at small separations (< 1 Mpc/h) in qualitative agreement with observations by Singh, Mandelbaum & More (2015).

($M_r < -22.6$) presented in this work. We show this agreement in Figure 16. The gray curves show their power-law fits on small scales, and their NLA model fits across all scales (see their Table 1). Given that Tenneti et al. (2015) match the observed alignments of LRGs with their simulation data, it is expected that Horizon-AGN will equally match observations if the redshift dependence of the signal is similar to that found in that work. However, we find that their fits significantly overestimate the alignment amplitude of the whole galaxy sample presented in Figure 15, and this could indicate a steeper luminosity dependence of alignments in Horizon-AGN. Joachimi et al. (2011) indeed found a steeper luminosity dependence of observed LRG alignments compared to Tenneti et al. (2015). However, such a steep luminosity dependence is not sufficient to reproduce the alignment amplitude of Figure 15. It is possible that blue galaxy alignments are suppressing our results in that figure as well.

Codis et al. (2014) found no alignments for red galaxies using Horizon-AGN at $z = 1.2$. In that work, spin was used as a proxy for galaxy shape. Our results are consistent with theirs. We have shown that for red galaxies, the spin alignment signal is very different from the shape alignment signal, as these galaxies do not carry significant angular momentum.

Velliscig et al. (2015) studied the galaxy-halo misalignment comparing the shapes of stars and hot gas to that of the underlying DM halo using the EAGLE smoothed-particle-hydrodynamics simulation. They found that the alignment of the stellar component with the entire DM halo increases

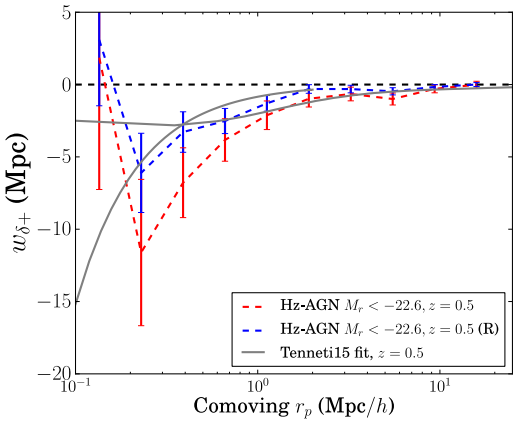


Figure 16. Measurement of $w_{\delta+}$ from Horizon-AGN for the most luminous galaxies, with absolute r-band magnitudes $M_r < -22.6$ ($\sim$ 800 galaxies). The blue curve corresponds to shapes measured using the reduced inertia tensor and the red curve, using the simple inertia tensor. We also show alignment model fits from Tenneti et al. (2015) (gray), which are in good agreement with our results. Tenneti et al. (2015) fit a power-law on small scales ($0.1-1$ Mpc/h, shown here in the range $0.1-2$ Mpc/h) and the NLA model on large scales ($6-25$ Mpc/h, extrapolated here over all scales shown). While the agreement is good at high luminosity, we find that the fits by Tenneti et al. (2015) significantly overpredict the alignment signal for the whole sample shown in Figure 15.

as a function of distance from the center to the subhalo and as a function of halo mass. This is in qualitative agreement with our finding that the shape alignment signal is reduced when using the reduced inertia tensor, which puts more weights towards the inner regions of galaxy. However, they also find that misalignment angles between the stellar and the DM components are larger for early-type than for late-type galaxies using the simple inertia tensor. This result will require further comparison, as the DM is expected to have stronger alignments than the baryons (Okumura, Jing & Li 2009) and given that we find a stronger shape alignment signal in Horizon-AGN for early-types than for late-types.

7 CONCLUSIONS

We have studied the alignments of galaxies, as traced by their stellar particles, using the Horizon-AGN simulation.

The main result of this paper is the clear identification of two different alignment mechanisms for disc-like galaxies and spheroidals, in qualitative agreement with theoretical expectations (Catelan, Kamionkowski & Blandford 2001; Codis, Pichon & Pogosyan 2015). This is the first time that these two mechanisms are clearly separated in a hydrodynamical cosmological simulation. In contrast, previous work by Tenneti et al. (2014) was unable to distinguish between red/blue galaxy orientation-separation alignments. This is likely a consequence of the different methods used to solve for the hydrodynamics (AMR in Horizon-AGN compared to smoothed particle hydrodynamics in MassiveBlack II), which result in different galaxy properties and their evolution with redshift.

We also reached the following conclusions:

- High mass (low V/σ) galaxies are elongated pointing towards other galaxies. This trend is preserved when projected correlations of the density field and galaxy shapes are considered.
- There is a preferential tangential orientation of disc-like galaxies around spheroidals. This trend is diluted in projection, possibly due to the equal weighting of the different Π bins in Equation (8) and the reduced contribution of face-on discs to Equation (7). This suggests that in order to extract the maximum possible information from intrinsic alignments in simulations, it could be beneficial to avoid performing projections along the line of sight, and rather access the full three dimensional information on alignments provided by the simulations. This would increase the signal-to-noise ratio in intrinsic alignments constraints from simulations and allow for more accurate forecasts of intrinsic alignment contamination to weak lensing in future surveys, such as *Euclid*[6] (Laureijs et al. 2011), the Large Synoptic Survey Telescope[7] (Ivezic et al. 2008) and WFIRST[8] (Green et al. 2011).
- We are able to describe w_{g+} across all scales probed and for all galaxies in the simulation with $> 10^9\,\mathrm{M}_\odot$ using a power-law in the case where shapes are obtained with the simple inertia tensor. The NLA and LA model tend to underestimate the power on small scales; a conclusion also reached by Singh, Mandelbaum & More (2015) using low redshift observations of LRGs. Fits to w_{g+} in the reduced inertia tensor case result in an alignment amplitude that is consistent with null for our complete sample of galaxies. This does not imply that alignments are not potential contaminants to weak lensing measurements. It will be necessary to match the shape measurement and galaxy selection done in observations to make more quantitative assessments of contamination from alignment to future surveys. The alignments of the most luminous galaxies in Horizon-AGN are, in fact, in agreement with work by Tenneti et al. (2015) and the alignments of bright LRGs in SDSS.
- Galactic kinematics, as quantified by V/σ, is a good proxy for the level of alignment. We also find that, given the existing correlation between V/σ and $u-r$ colour, the latter can also be used as proxy to separate galaxy populations with different sensitivity to alignment. On the contrary, the amplitude of the intrinsic alignment signal is not monotonic with stellar mass, which we interpret as a consequence of the wide distribution of stellar masses at low V/σ.
- We emphasise that correlations of spins (or shapes) and separation are not contaminated by grid-locking because this effect averages to null for position-shape correlations (as shown in Appendix A). This is not necessarily true for two-point auto-correlations of spins and/or shapes, and so we refrain from giving these a physical interpretation in the main body of the manuscript, although we present them in Appendix B for completeness.
- The choice of shape estimator can have a large impact on predictions for intrinsic alignments from cosmological simulations. To make accurate predictions of intrinsic alignment impact on weak lensing surveys, we expect that it will be necessary to create mock images of simulated galaxies in a manner that takes into account photometric depth, noise and convolution by the point-spread function. We defer this and a study of redshift evolution and selection cuts on the galaxy sample to future work.

ACKNOWLEDGMENTS

This work has made use of the HPC resources of CINES (Jade and Occigen supercomputer) under the time allocations 2013047012, 2014047012 and 2015047012 made by GENCI. This work is partially supported by the Spin(e) grants ANR-13-BS05-0005 (`http://cosmicorigin.org`) of the French *Agence Nationale de la Recherche* and by the ILP LABEX (under reference ANR-10-LABX-63 and ANR-11-IDEX-0004-02). Part of the analysis of the simulation was performed on the DiRAC facility jointly funded by STFC, BIS and the University of Oxford. NEC is supported by a Beecroft Postdoctoral Research Fellowship. We thank S. Rouberol for running smoothly the `Horizon` cluster for us. NEC thanks Rachel Mandelbaum for useful discussion regarding the grid-locking effect.

[6] `http://sci.esa.int/euclid`
[7] `http://www.lsst.org`
[8] `http://wfirst.gsfc.nasa.gov/`

References

Andrae R., Jahnke K., 2011, *MNRAS*, 418, 2014

Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007, *ApJ Let.*, 655, L5
Aubert D., Pichon C., Colombi S., 2004, *MNRAS*, 352, 376
Bailin J., Steinmetz M., 2005, *ApJ*, 627, 647
Bernstein G. M., Jarvis M., 2002, *AJ*, 123, 583
Blazek J., McQuinn M., Seljak U., 2011, *JCAP*, 5, 10
Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
Bridle S., King L., 2007, New Journal of Physics, 9, 444
Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002, *MNRAS*, 333, 501
Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
Catelan P., Kamionkowski M., Blandford R. D., 2001, *MNRAS*, 320, L7
Chisari N. E., Dunkley J., Miller L., Allison R., 2015, ArXiv e-prints
Chisari N. E., Dvorkin C., 2013, *JCAP*, 12, 29
Chisari N. E., Dvorkin C., Schmidt F., 2014, *Phys. Rev. D*, 90, 043527
Codis S., Dubois Y., Pichon C., Devriendt J., Slyz A., 2014, ArXiv e-prints
Codis S. et al., 2015, *MNRAS*, 448, 3391
Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, *MNRAS*, 427, 3320
Codis S., Pichon C., Pogosyan D., 2015, ArXiv e-prints
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, *MNRAS*, 420, 2662
Dubois Y. et al., 2014, *MNRAS*, 444, 1453
Dubois Y., Teyssier R., 2008, *A&A*, 477, 79
Green J. et al., 2011, ArXiv e-prints
Greggio L., Renzini A., 1983, *A&A*, 118, 217
Gunn J. E. et al., 2006, *AJ*, 131, 2332
Haardt F., Madau P., 1996, *ApJ*, 461, 20
Hahn O., Carollo C. M., Porciani C., Dekel A., 2007, *MNRAS*, 381, 41
Hahn O., Teyssier R., Carollo C. M., 2010, *MNRAS*, 405, 274
Heymans C. et al., 2013, *MNRAS*, 432, 2433
Hirata C. M., Mandelbaum R., Ishak M., Seljak U., Nichol R., Pimbblet K. A., Ross N. P., Wake D., 2007, *MNRAS*, 381, 1197
Hirata C. M. et al., 2004, *MNRAS*, 353, 529
Hirata C. M., Seljak U., 2004, *Phys. Rev. D*, 70, 063526
Hirata C. M., Seljak U., 2010, *Phys. Rev. D*, 82, 049901
Huff E. M., Krause E., Eifler T., George M. R., Schlegel D., 2013, ArXiv e-prints
Hui L., Zhang J., 2002, ArXiv Astrophysics e-prints
Hung C.-L., Ebeling H., 2012, *MNRAS*, 421, 3229
Hutsemékers D., Braibant L., Pelgrims V., Sluse D., 2014, *A&A*, 572, A18
Ivezic Z. et al., 2008, ArXiv e-prints
Joachimi B., Bridle S. L., 2010, *A&A*, 523, A1
Joachimi B. et al., 2015, ArXiv e-prints
Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, *A&A*, 527, A26
Joachimi B., Schneider P., 2010, ArXiv e-prints
Jones B. J. T., van de Weygaert R., Aragón-Calvo M. A., 2010, *MNRAS*, 408, 897
Kaiser N., 1987, *MNRAS*, 227, 1
Kennicutt, Jr. R. C., 1998, *ApJ*, 498, 541
Kiessling A. et al., 2015, ArXiv e-prints
Kirk D. et al., 2015, ArXiv e-prints
Kirk D., Rassat A., Host O., Bridle S., 2012, *MNRAS*, 424, 1647
Komatsu E., Smith K. M., Dunkley J., et al., 2011, *ApJ Sup.*, 192, 18
Krause E., Eifler T., Blazek J., 2015, ArXiv e-prints
Krumholz M. R., Tan J. C., 2007, *ApJ*, 654, 304
Laigle C. et al., 2015, *MNRAS*, 446, 2744
Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
Laureijs R. et al., 2011, ArXiv e-prints
Leitherer C., Ortiz Otálvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, *ApJ Sup.*, 189, 309
Leitherer C., Schaerer D., Goldader J. D., et al., 1999, *ApJ Sup.*, 123, 3
Libeskind N. I., Hoffman Y., Knebe A., Steinmetz M., Gottlöber S., Metuki O., Yepes G., 2012, *MNRAS*, 421, L137
Mandelbaum R. et al., 2011, *MNRAS*, 410, 844
Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006a, *MNRAS*, 367, 611
Mandelbaum R., Seljak U., Cool R. J., Blanton M., Hirata C. M., Brinkmann J., 2006b, *MNRAS*, 372, 758
Okumura T., Jing Y. P., Li C., 2009, *ApJ*, 694, 214
Paz D. J., Stasyszyn F., Padilla N. D., 2008, *MNRAS*, 389, 1127
Pen U.-L., Lee J., Seljak U., 2000, *ApJ Let.*, 543, L107
Rasera Y., Teyssier R., 2006, *A&A*, 445, 1
Salpeter E. E., 1955, *ApJ*, 121, 161
Schaefer B. M., 2009, International Journal of Modern Physics D, 18, 173
Schaefer B. M., Merkel P. M., 2015, ArXiv e-prints
Schmidt F., Chisari N. E., Dvorkin C., 2015, ArXiv e-prints
Schneider M. D., Bridle S., 2010, *MNRAS*, 402, 2127
Schneider M. D., Frenk C. S., Cole S., 2012, *JCAP*, 5, 30
Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
Singh S., Mandelbaum R., More S., 2015, *MNRAS*, 450, 2195
Slosar A., White M., 2009, *JCAP*, 6, 9
Sousbie T., Pichon C., Colombi S., Pogosyan D., 2008, *MNRAS*, 383, 1655
Sutherland R. S., Dopita M. A., 1993, *ApJ Sup.*, 88, 253
Tempel E., Libeskind N. I., 2013, *ApJ Let.*, 775, L42
Tenneti A., Mandelbaum R., Di Matteo T., Feng Y., Khandai N., 2014, *MNRAS*, 441, 470
Tenneti A., Singh S., Mandelbaum R., Matteo T. D., Feng Y., Khandai N., 2015, *MNRAS*, 448, 3522
Teyssier R., 2002, *A&A*, 385, 337
Troxel M. A., Ishak M., 2012, *MNRAS*, 419, 1804
Troxel M. A., Ishak M., 2015, *Phys. Rep.*, 558, 1
Velliscig M. et al., 2015, ArXiv e-prints
Zhang P., 2010, *MNRAS*, 406, L95
Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, *ApJ*, 706, 747

APPENDIX A: GRID-LOCKING

Adaptive-mesh-refinement simulations are thought to be subject to grid-locking systematics, whereby the galaxy spins become aligned with the directions of the grid. We have verified that this effect is indeed present in the

Horizon-AGN simulation, but we demonstrate in this appendix that it does not affect position-shape or position-spin correlations.

A1 Correlations with the box

We consider possible systematics that might arise from correlations of the galaxy orientations with preferential directions of the simulation box. Figure A1 shows the excess fraction of galaxies as a function of $|\cos(\mathbf{s}\cdot\{\mathbf{x},\mathbf{y},\mathbf{z}\})|$, where $\mathbf{s}$ is the galaxy spin and we average over x, y, z. There is a clear excess of galaxies at directions perpendicular and parallel/anti-parallel to the grid, and a decrement at intermediate angles. A similar behaviour is observed for the direction minor axes in the same figure. The higher mass galaxies are less affected, but the trend is not monotonic with mass. We have also considered selecting the galaxy sample by median stellar age, resolution and colour. While mass and median stellar age cuts can help remove the systematics, it is not possible to define clean sample without removing the vast majority of the galaxies. Moreover, it is not clear how these selection cuts affect the intrinsic alignment measurement and its comparison to current observational constraints. For these reasons, we decide to avoid placing cuts on the galaxy sample to reduce the grid-locking effects. Instead, we show in the next section that the correlation between galaxy shapes, spins and the box have no impact on the $w_{\delta+}$ statistic presented in Section 4.

We have also considered the grid-locking signal of shapes obtained through the reduced inertia tensor. In this case, we find that the grid-locking signal is smaller than for Figure A1.

A2 Contamination to intrinsic alignments

Grid-locking creates correlations of the spins and shapes of galaxies with $\{x, y, z\}$. However, due to the periodic boundary conditions of the box, we expect that the grid-locking will average to null around any arbitrary point within the box. We test this hypothesis by randomizing the position of galaxies while computing the relative orientation of their spins and shapes around these random positions. Figure A2 shows an example correlation for shapes defined from the orientation of the minor axis and galaxies in three mass bins. We do not find significant correlations when positions are randomized. Notice that this test is sufficient only under the assumption that the large-scale structure is not itself locked to the grid. This assumption was shown to be valid for the dark matter large-scale structure in Horizon-AGN (Dubois et al. 2014). We further test this assumption for galaxies in Horizon-AGN in Figure A3, where we show the excess probability of spin alignments (top panel) and separation vectors alignments (bottom panel) with the grid for pairs of galaxies separated by less than 25 Mpc/h in the simulation. Spins are clearly correlated with the box axes, as demonstrated in Fig. A1, while this is not the case of the separation vectors.

We apply a similar procedure to confirm that grid-locking does not contribute to $w_{\delta+}$. We obtain the projected correlation function of the random positions and the + (w_{r+}) and × components of the shapes. Both of these correlations are consistent with null. This confirms that spherical averages remove the effect of grid-locking on position-shape correlations. Figure A4 shows w_{r+} for five mass bins as a function of r_p.

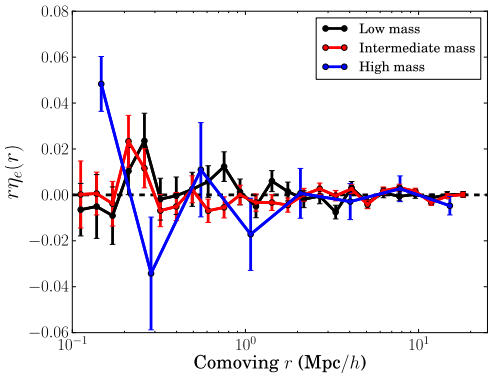


Figure A2. Correlation between minor axis orientation and separation vector as a function of separation using randomized positions. Black lines correspond to low mass galaxies ($10^9\,\mathrm{M}_\odot < M_* < 10^{9.5}\,\mathrm{M}_\odot$); red lines, to intermediate mass galaxies ($10^{9.5}\,\mathrm{M}_\odot < M_* < 10^{10.6}\,\mathrm{M}_\odot$); and blue lines, to high mass galaxies ($M_* > 10^{10.6}\,\mathrm{M}_\odot$).

Notice that this procedure is not applicable to two-point auto-correlations of the spin or the minor axes. In other words, the impact of grid-locking on auto-correlations cannot be fully quantified by randomizing galaxy positions. This procedure would only measure the impact of uniform correlations of galaxy shapes or spins across $\{x, y, z\}$, but two galaxies clustered together could be increasingly affected by grid-locking. We present auto-correlations in Appendix B, along with possible interpretations of the signals as a result of physical alignments or grid-locking.

APPENDIX B: AUTO-CORRELATIONS

We have so far shown that the spins and shapes of galaxies tend to align with the grid, and that these grid-locking effects do not affect our measurements of position-shape (or position-spin) correlations. Unfortunately, the symmetry arguments that apply to position-shape correlations no longer hold when auto-correlations are considered. As a consequence, it is harder to separate the physical alignment signal and the grid-locking from auto-correlation of galaxies spins or shapes. In this section, we present those auto-correlations and we give the two possible interpretations of the signal.

The spin-spin (SS) correlation function measures the relative orientation of the angular momenta of two galaxies separated by a comoving distance vector $\mathbf{r}$,

$$\eta_S(r) = \langle|\hat{\mathbf{s}}(\mathbf{x})\cdot\hat{\mathbf{s}}(\mathbf{x}+\mathbf{r})|^2\rangle - 1/3\,, \tag{B1}$$

and the ellipticity-ellipticity (EE) correlation function is given by

$$\eta_E(r) = \langle|\hat{\mathbf{e}}(\mathbf{x})\cdot\hat{\mathbf{e}}(\mathbf{x}+\mathbf{r})|^2\rangle - 1/3\,, \tag{B2}$$

where $\hat{\mathbf{e}}$ is the direction of the minor axis. Similarly to ξ_{g+}, defined in Equation (6), we can define the auto-correlation

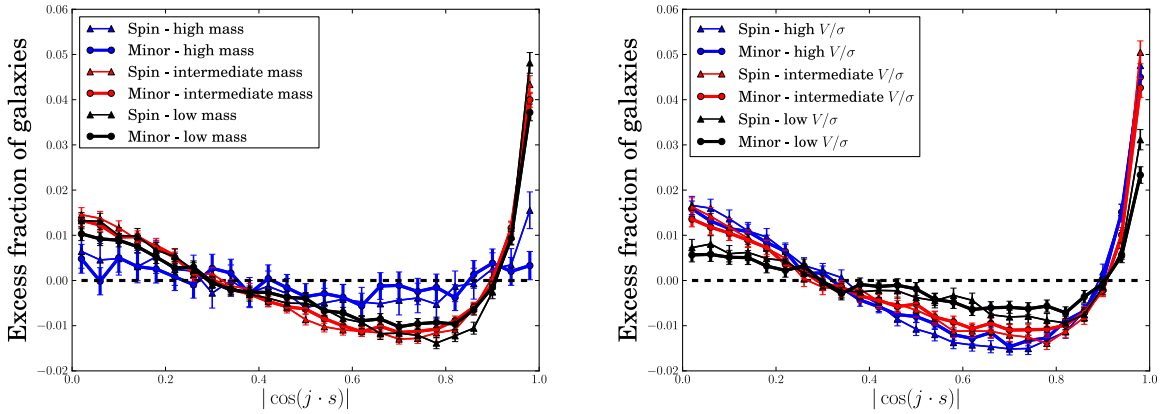


Figure A1. Excess fraction of galaxies with respect to a uniform distribution as a function of the angle between the spin/minor axis with the directions of the box (averaged over $j = \{x, y, z\}$). In the left panels, the galaxy sample is split by mass in the following bins: $10^9\,\mathrm{M}_\odot < M_* < 10^{9.5}\,\mathrm{M}_\odot$ (black), $10^{9.5}\,\mathrm{M}_\odot < M_* < 10^{10.6}\,\mathrm{M}_\odot$ (red) and $M_* > 10^{10.6}\,\mathrm{M}_\odot$ (blue). In the left panel, the galaxy sample is split by V/σ: $V/\sigma < 0.55$ (black), $0.55 < V/\sigma < 0.79$ (red) and $V/\sigma > 0.79$ (blue). There is an excess of galaxies in directions parallel and perpendicular to the grid axes, which is less significant for high mass and low V/σ galaxies. This effect averages to null for the angular and projected correlations of spin/minor axis and separation, but it could give rise to a non-negligible two-point correlation.

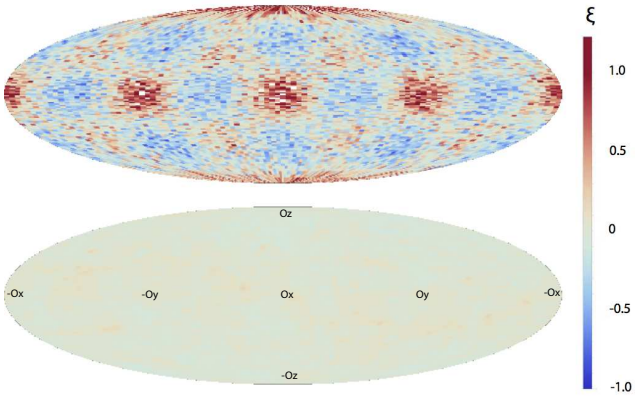


Figure A3. Distribution of spins (top panel) and separation vectors of galaxy pairs separated by less than 25 Mpc/h (bottom panel) on the sphere. The excess probability ξ defined so that the PDF reads $P(\cos\theta, \phi) = (1 + \xi)/4\pi$ is colour-coded from dark blue (-1) to dark red (+1). If the spins are clearly correlated with the box axes, this is not the case of the separation vectors for which the magnitude of the excess probability is smaller (lighter colours) and the pattern does not seem to be correlated with the grid.

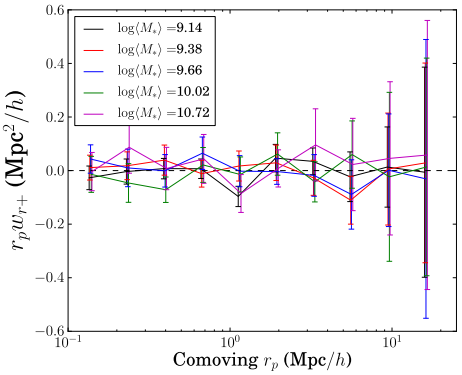


Figure A4. Projected correlation function between random positions and + component of the ellipticity. This result is consistent with null, which suggests that any systematics coming from grid-locking is not affecting our $\delta+$ or $\delta\times$ measurements. The different colours represent different bins in stellar mass and the points are slightly displaced to higher r_p for each mass bin for visual clarity. The mean stellar mass in each bin is indicated in the legend.

function of + components of the ellipticity

$$\xi_{++}(r_p, \Pi) = \frac{S_+ S_+}{RR}\,, \tag{B3}$$

and its projection along the line-of-sight, w_{++}; and analogously for $\xi_{\times\times}$ and $w_{\times\times}$.

Figure B1 shows the relative orientation of the spins of two galaxies (right panel) separated by a comoving distance of r. The correlation is significant for low and intermediate V/σ galaxies. For pair separations $\lesssim 10\,h^{-1}$ Mpc, the spins tend to align with each other and the alignment decreases on larger scales. This trend is monotonic with V/σ. The results for the auto-correlation between minor axes or between spins are very similar when minor axes are computed using the reduced inertia tensor (middle panel of Figure B1). The signal slightly decreases when the simple inertia tensor is used (left panel). These three dimensional auto-correlations of galaxy orientations seem to be dominated by the contribution of disc-like galaxies. In Section 5.1, we saw that disc galaxies tend orient their spin (also reduced minor axis) tangentially around over-densities. Figure B1 is qualitatively compatible with those results. Similarly, we find no auto-correlation between the shapes of low V/σ galaxies in the left panel of Figure B1.

We show the orientation-separation correlations as a function of mass in Figure B2. Intermediate and low mass galaxies tend to have a stronger auto-correlation than high mass galaxies. The signal is stronger for interme-

diate mass galaxies, which is expected both from grid-locking and from the physical signal due to the contribution of the high V/σ population. Similar qualitative results were obtained by Codis et al. (2015) at $z = 1.2$ using the Horizon-AGN simulation with the same mass selection (see their Figure 8 for a direct comparison). The trend with mass is similar, but the amplitude of the signal is higher in our case, suggesting that the spin auto-correlation increases at lower redshift and at fixed mass. Codis et al. (2015) also found that separating the galaxy population by $u-r$ colour into red and blue galaxies resulted in a significant spin alignment signal for blue galaxies, while no spin alignment was found for red galaxies. This is also in agreement with our results, given the strong correlation between colour and V/σ.

The projected auto-correlation functions of galaxy shapes in three mass bins and three V/σ are shown in Figure B3 decomposed into E-modes $(++)$ and B-modes $(\times\times)$. Qualitatively, we observe that the projected auto-correlation of shapes is more significant for intermediate mass galaxies and high V/σ. This is consequence of a selection effect: galaxies with intermediate masses also tend to have higher V/σ in Figure 1. We also find that the amplitude of the auto-correlation decreases when the reduced inertia tensor is used. We attribute this to the fact that w_{++} and $w_{\times\times}$ carry a double weighting by galaxy ellipticities, which are also more round in the reduced case.

Both E-modes and B-modes are present with similar amplitudes and scale-dependence in the top middle and right bottom panels of Figure B3. Using the diagonals of the jack-knife covariance, we find that the signal is different from null at the 89% C.L. (85% C.L.) for the E-modes (B-modes) of intermediate mass galaxies when using the simple inertia tensor; and higher (99% C.L.) using the reduced inertia tensor. For high V/σ galaxies, the signal is different from null at the 79% C.L. for E-modes and 81% for B-modes using the simple inertia tensor; and higher ($> 93\%$ C.L.) using the reduced inertia tensor. All other bins of mass and V/σ have auto-correlations generally consistent with being null at the 2σ level.

Overall, we find that auto-correlations of galaxy shapes or spins are dominated by the contribution of disc-like galaxies. We find no contribution from spheroidals to these auto-correlations. Moreover, we find equal contribution of E-modes and B-modes in the projected correlation functions. If the signal is to be interpreted as a physical signal, the presence of an auto-correlation of disc-like tracers is expected, and the trend with mass/dynamics agrees with that found in Section 5.1.

On the other hand, the absence of an auto-correlation for the shapes of spheroidals seems inconsistent with the presence of auto-correlations for discs. This could suggest that at least part of this signal is produced by grid-locking. E-mode and B-mode auto-correlations could also be produced by both a physical signal or by grid-locking. In the latter case, B-modes would be a consequence of preferential alignments of the simulated galaxies with the diagonals of the grid. Ideally, two-point correlations such as those presented in Figures B1 and B3 could potentially be used to constrain a simulation-dependent model of grid-locking. However, there is not sufficient evidence that current intrinsic alignment models predict the correct relation between density-shape and shape-shape correlations, especially on such small scales as probed in this work. Hence separating the grid-locking from the physical alignment signal in auto-correlations remains a difficult task for adaptive-mesh-refinement codes like ramses.

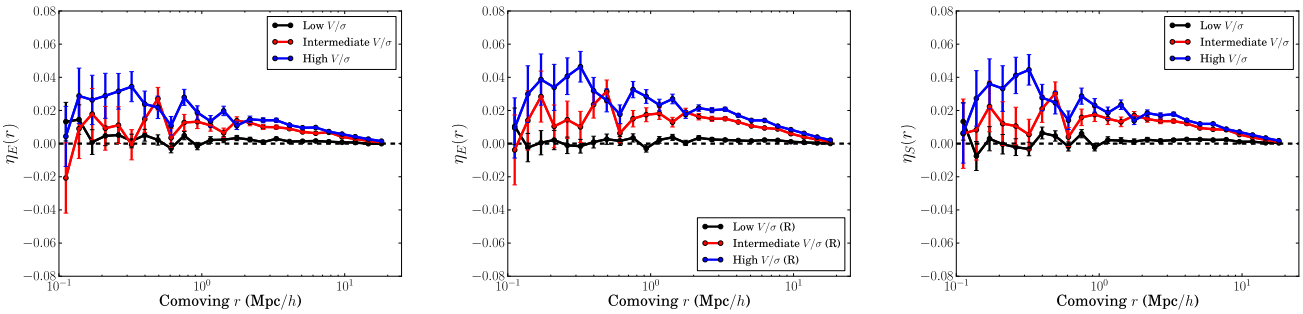


Figure B1. Correlations between minor axes obtained from the simple inertia tensor (left panel), from the reduced inertia tensor (middle panel) and between spins (right panel) as a function of comoving separation. Black lines correspond to low V/σ galaxies ($V/\sigma < 0.55$); red lines, to intermediate V/σ galaxies ($0.55 < V/\sigma < 0.79$); and blue lines, to high V/σ galaxies ($V/\sigma > 0.79$).

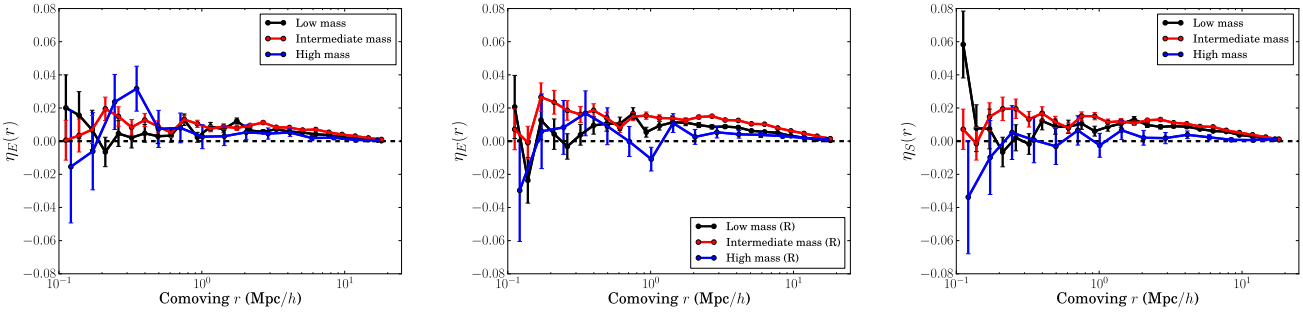


Figure B2. Correlations between minor axes obtained from the simple inertia tensor (left panel), from the reduced inertia tensor (middle panel) and between spins (right panel) as a function of comoving separation. Black lines correspond to low mass galaxies ($10^9\,\mathrm{M}_\odot < M_* < 10^{9.5}\,\mathrm{M}_\odot$); red lines, to intermediate mass galaxies ($10^{9.5}\,\mathrm{M}_\odot < M_* < 10^{10.6}\,\mathrm{M}_\odot$); and blue lines, to high mass galaxies ($M_* > 10^{10.6}\,\mathrm{M}_\odot$).

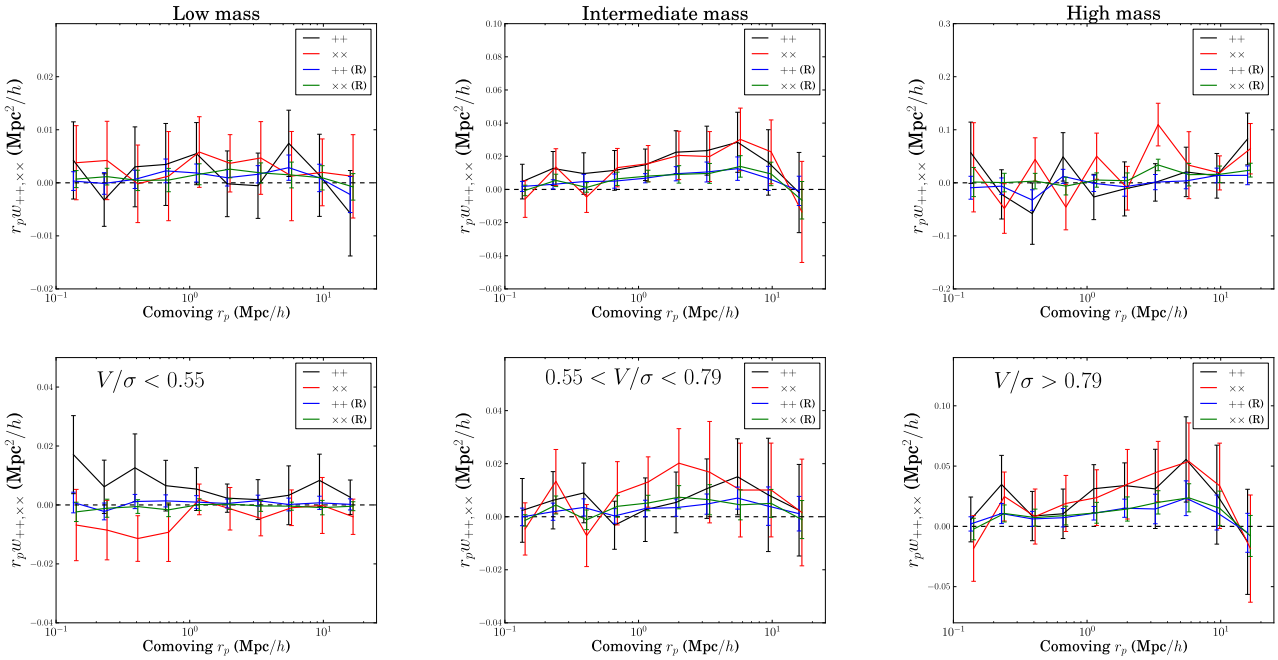


Figure B3. Projected auto-correlation functions of + and × components of the shape as a function of projected radius in three mass bins (upper panels) and three V/σ bins (lower panels). In all panels, the ++ correlation is shown in black for shapes measured using the simple inertia tensor; correspondingly, the ×× correlation is shown in red. All panels show the impact of using the reduced inertia tensor (blue lines for ++ and green lines for ××). All ×× correlations are arbitrarily displaced to 5% higher r_p for visual clarity.

5 Alternative probes of cosmology using the large-scale structure of the Universe

In this chapter, I will present two alternative probes of cosmology beyond two-point correlation fonctions : topological invariants of the density field and count-in-cells statistics. In a first part, I will show how to predict topological invariants for Gaussian random fields in real space. Then, I will extend this theory to weakly non-Gaussian fields and to redshift space where some level of symmetry is broken. This formalism gives access to $D(z)$ and to $f = \mathrm{d}\log D/\mathrm{d}\log a$- the linear growth rate of structures- which probe in particular the equation of state of dark energy and possible modifications of gravity. In a second part, I will exhibit some recent results in the field of count-in-cells statistics. The idea is to measure the mean density within concentric spheres and study their joint statistics. Owing to the spherical symmetry, one can do a mathematical conjecture that leads to surprisingly accurate predictions even in the mildly non-linear regime where standard perturbation theory calculations break down. Implications for cosmological measurements with future surveys like Euclid will be discussed.

Publications

5.1 Overview

In modern cosmology, random fields (3D or 2D) are fundamental ingredients in the description of the large-scale matter density and the Cosmic Microwave Background (CMB). The large-scale structure of the matter distribution in the Universe is believed to be the result of the gravitational growth of primordial nearly-Gaussian small perturbations originating from quantum fluctuations. Deviations from Gaussianity inevitably arise due to the non-linear dynamics of the growing structures, but may also be present at small, but potentially detectable levels in the initial seed inhomogeneities. Thus, studying non-Gaussian signatures in the random fields of cosmological data provides methods to learn both the details of early Universe's physics and mechanisms for structure's growth, addressing issues such as the matter content of the Universe, the role of bias between galaxies and dark matter distributions, and whether it is dark energy or a modification to Einstein's gravity that is responsible for the acceleration of the Universe's expansion.

With the advent of large galaxy surveys (e.g. SDSS and in the coming years Euclid, LSST), astronomers have ventured into the era of statistical cosmology and big data. Hence, there is a dire need for them to build tools that can efficiently extract as much information as possible from these huge data sets at high and low redshift. In particular, this means being able to probe the non-linear regime of structure formation. The most commonly used tools to extract statistical information from the observed galaxy distribution are N-point correlation functions which quantify how galaxies are clustered. In our initially Gaussian Universe the matter density field is fully described by its two-point correlation function. However departure from Gaussianity occurs when the growth of structure becomes non-linear (at later times or smaller scales), providing information that is not captured by the two-point correlation function but is recorded in part in the three-point correlation function. Obviously N-point correlation functions are increasingly difficult to measure when N increases. They are noisy, subject to cosmic variance and highly sensitive to systematics such as the complex geometry of surveys. From a theoretical point of view, N-point statistics are also increasingly difficult to predict in the context of cosmological perturbation theory. Figure 5.1 sums up in a table the observables that have been predicted so far using perturbation theory. Typically, the power spectrum at two-loop order, the bispectrum at one-loop in some specific configurations and the trispectrum at tree order are within reach from standard computations but it is highly difficult to go deeper in the non-linear regime and to higher order statistics.

It is thus essential to find alternative estimators to extract information from the non-linear regime of structure formation in order to complement these classical probes. This is in particular critical if we are to understand the origin of dark energy, which accounts for ~ 70 % of the energy budget of our Universe! In this section, we propose to focus on two such kinds of very promising cosmological observables : topological and geometrical estimators in section 5.1.1 and count-in-cell statistics in section 5.1.2.

5.1.1 Topology and geometry of random fields

5.1.1.1 Minkowski functionnals

The first of these two promising probes involves the topological features of the density field. The topology of a 3D field can be described by only four functionals –named after Minkowski–

	leading order LO	order 1 NLO	order 2 NNLO	order 2.5	order 3	... order p
2-point statistics	OK	OK	OK	EFT	partial exact results	partial resummations
3-point statistics	OK	OK (but not systematics				partial resummations
4-point statistics	OK					
N-point statistics	OK, for **topological invariants** OK, in specific geometries **(count-in-cell)**					

Figure 5.1: State-of-the-art predictions of N-point statistics at different orders in perturbation theory. Standard calculations have only been carried out for small N and at low order (light blue). Nevertheless, alternative observables (light orange) can be found such as topological invariants (that are equivalent to a combination of standard PT results but are supposedly more robust) and count-in-cell statistics (that can be predicted for higher non-Gaussianities because of the spherical symmetry). They are the prime focus of this section of the manuscript.

that can be analytically computed when non-Gaussianities are weak enough (see section 2.2.2). On top of these four topological invariants, it is also of interest to study geometrical estimators such as extrema counts or the statistical properties of the skeleton (length of filaments, ...). The advantage of these estimators compared to the usual use of N-point correlation functions and their counterpart in Fourier space is twofold. First, they are bias-independent in the sense that they are invariant under any monotonic transformation. This property is clearly seen for extrema counts as the number of extrema of a field f or of any monotonic transform of this field is trivially the same (the value of each extremum changes but not its existence). On top of that, the topological features should by essence be more robust i.e. easier to measure in real data because for instance less sensitive to masks or noise.

In [Codis et al., 2013] (see paper 5.2), I developed the theory of these estimators to all order in non-Gaussianity. I also showed how redshift space distortion can be accounted for within this context. When predicting these estimators, one gets a function of the moments of the field and its derivative. Perturbation theory then predicts these moments of the field and its derivatives as a function of the underlying cosmology and in particular as a function of the variance σ and the redshift distortion parameter β. With upcoming 3D spectroscopic surveys such as Euclid, the statistical analysis of the topology of our redshift-distorted Universe will therefore allow us to robustly measure weighted moments of the multi-spectra as a function of redshift, and henceforth quantify the cosmic evolution of the equation of state of dark energy and possible departure from General Relativity. In [Codis et al., 2013], I specifically investigated how topological estimators in redshift space can be used to estimate the cosmic evolution of the growth of structure (the so-called β parameter). Let me therefore start by connecting the statistics of non gaussian critical sets to the underlying cumulants of the field. I will then move on to the growth rate $D(z)$ and finally on how to constrain the equation of state of dark energy

with these statistics.

From Non-Gaussian critical sets to cumulants Let us first think about the statistical properties of critical sets such as those described in section 2.2.2. For instance the Euler characteristic of density contours below a threshold in contrast, or the differential number counts of maxima within some contrast or the length of the skeleton above a threshold. Their statistical expectation requires the knowledge of the so-called joint distribution of the field and its one and second derivatives. Since these descriptors are stationary, they should only involve (partly)-isotropic combinations of the field variables. For instance, the field x, the modulus square of the rescaled gradient q^2 and the eigenvalues of the rescaled Hessian in real space λ_1, λ_2, λ_3. More invariants are needed in redshift space because there are more degrees of freedom (the invariance by rotation is replaced by an invariance by rotation on the sky). We choose combinations of these invariants that are as uncorrelated as possible in the Gaussian limit to simplify the integrals. We eventually end up with the following Gaussian joint PDF in two dimensions

$$G(x, q^2, \zeta, J_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}x^2 - q^2 - \frac{1}{2}\zeta^2 - J_2\right), \tag{5.1}$$

where $\zeta = (J_1 + \gamma x)/\sqrt{1-\gamma^2}$, $J_1 = \lambda_1 + \lambda_2$ and $J_2 = (\lambda_1 - \lambda_2)^2$. It has to be noted that two variables x and ζ are linear in the fields and therefore follow a Gaussian distribution whose associated orthogonal polynomials are the Hermite polynomials H_i and the other two variables q^2 and J_2 are quadratic in the fields and therefore follow an exponential distribution whose associated orthogonal polynomials are the Laguerre polynomials L_i. The Gaussian PDF in three dimensions can also be written in the same way as equation 5.1. In redshift space, similar PDFs are derived that now depend on more variables (6 instead of 4 in 2D and 8 instead of 5 in 3D). Those PDFs are explicitly written in paper 5.2. In order to describe weakly non-Gaussian fields (such as the evolving density field), a Gram-Charlier expansion of the non-Gaussian joint PDF can be used (see section 2.2.1.4) so that, in real 2D space, we get

$$P(x, q^2, \zeta, J_2) = G \times \left[1 + \sum_{n=3}^{\infty} \sum_{i+j+2k+2l=n} \frac{(-1)^{j+l}}{i!\,j!\,k!\,l!} \left\langle x^i q^{2k} \zeta^j J_2^l \right\rangle_{GC} H_i(x) L_k(q^2) H_j(\zeta) L_l(J_2)\right],$$

where $\left\langle x^i q^{2k} \zeta^j J_2^l \right\rangle_{GC} = (-1)^{j+l} j!\,l! \left\langle H_i(x) L_k(q^2) H_j(\zeta) L_l(J_2) \right\rangle$ are the Gram-Charlier coefficients of the expansion.

Once the joint PDF is known, the genus and other estimators can be obtained by marginalizing this joint PDF under some constraints defined in section 2.2.2. For instance, extrema counts and the Euler characteristic require a condition of zero gradient. After some pages of algebra , we get the prediction for those estimators to all orders in non-Gaussianity which depends only on the contrast and the cumulants of the underlying field through the Gram-Charlier coefficients. Note that redshift space distortions break some level of symmetry and therefore more cumulants are needed.

For illustration purpose, we propose here to describe the procedure for the particular case of $\mathcal{N}_3(\nu)$, the area (per unit volume) of a 3D isosurface of the density field in redshift space at level ν. To compute this functional, it is sufficient to consider the joint PDF of the field and its first derivatives:

$$P(x, q_\perp^2, x_3) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2} - q_\perp^2 - \frac{x_3^2}{2}\right) \left[1 + \sum_{n=3}^{\infty} \sum_{\sigma_n} \frac{(-1)^j}{i!\,j!\,k!} \left\langle x^i q_\perp^{2j} x_3^k \right\rangle_{\mathrm{GC}} H_i(x) L_j(q_\perp^2) H_k(x_3)\right],$$

where $q_\perp^2$ is the modulus of the gradient perpendicular to the line-of-sight and x_3 the component of the gradient along the line-of-sight. The area of 3D isosurfaces now is

$$\mathcal{N}_3(\nu) = \frac{1}{\sigma}\int \mathrm{d}q_\perp^2 \mathrm{d}x_3 P(\nu, q_\perp^2, x_3)\sqrt{\sigma_{1\perp}^2 q_\perp^2 + \sigma_{1\parallel}^2 x_3^2}\,. \tag{5.2}$$

Computing the integrals, we express the results using the anisotropy parameter

$$\beta_\sigma \equiv 1 - \sigma_{1\perp}^2/2\sigma_{1\parallel}^2 \tag{5.3}$$

that measures the difference between the rms values of the line-of-sight and perpendicular components of the gradient. In the case of isotropy, $\beta_\sigma = 0$. In anisotropic situations, β_σ is positive, spanning the range $0 < \beta_\sigma \leq 1$, when the field changes faster in the z-direction ($\sigma_{1\parallel} > \frac{1}{2}\sigma_{1\perp}$) as is the case, for instance, in the linear regime of redshift corrections where $\beta_\sigma = \frac{4}{5}\beta(1+3\beta/7)/(1+6\beta/5+3\beta^2/7)$, with $\beta = f/b_1$. When the line-of-sight variations are smaller than the perpendicular one ($\sigma_{1\parallel} < \frac{1}{2}\sigma_{1\perp}$), β_σ is negative, $-\infty \leq \beta_\sigma < 0$, as is the case in the non-linear "finger of God" regime.

At Gaussian order, the expression (5.2) yields the result consistent with [Matsubara, 1996]

$$\mathcal{N}_3^{(0)}(\nu) = \frac{2}{\pi}\frac{\sigma_1}{\sqrt{3}\sigma}\frac{1-A(\beta_\sigma)}{\sqrt{1-2\beta_\sigma/3}}e^{-\nu^2/2}, \qquad A(\beta_\sigma) \equiv \frac{1}{2}\left(\beta_\sigma - T\left(\beta_\sigma\right) + \beta_\sigma T\left(\beta_\sigma\right)\right), \tag{5.4}$$

where the function T is defined as $T(\beta_\sigma) = 1/\sqrt{\beta_\sigma}\tanh^{-1}\left(\sqrt{\beta_\sigma}\right) - 1$ for $\beta_\sigma \geq 0$ and $T(\beta_\sigma) = 1/\sqrt{|\beta_\sigma|}\tan^{-1}\left(\sqrt{|\beta_\sigma|}\right) - 1$ for $\beta_\sigma < 0$. Under this definition $A(\beta_\sigma)$ describes a $\sim \beta_\sigma/3 + \beta_\sigma^2/15 + \ldots$ correction at small anisotropy $\beta_\sigma \to 0$. We see that in the Gaussian limit, the anisotropy has a very little effect on $\mathcal{N}_3$. The amplitude deviates from unity by less that 1% in the range $-1 < \beta_\sigma < 0.5$, as its series expansion $\propto 1 - \beta_\sigma^2/90\ldots$ attests. Even at extreme anisotropies, it changes only to ≈ 0.92 at $\beta_\sigma \to -\infty$ and ≈ 0.87 at $\beta_\sigma = 1$.

The computation of the $n = 3$ term corresponding to the first non-Gaussian correction can also be carried out

$$\begin{aligned}
N_3^{(1)}(\nu) &= \frac{e^{-\nu^2/2}}{2\pi\sigma}\sum_{i,j,k=0}^{i+2j+2k=3}\frac{(-1)^j}{i!\,j!\,(2k)!}\left\langle x^i q_\perp^{2j} x_3^{2k}\right\rangle_{\mathrm{GC}} H_i(\nu) \\
&\quad \times \int \mathrm{d}q_\perp^2 \mathrm{d}x_3 \exp\left(-q_\perp^2 - \frac{x_3^2}{2}\right)\sqrt{\sigma_{1\perp}^2 q_\perp^2 + \sigma_{1\parallel}^2 x_3^2}L_j(q_\perp^2)H_{2k}(x_3) \\
&= \frac{e^{-\nu^2/2}}{2\pi\sigma}\left[\frac{1}{3!}\left\langle x^3\right\rangle_c H_3(\nu)\int \mathrm{d}q_\perp^2 \mathrm{d}x_3 \exp\left(-q_\perp^2 - \frac{x_3^2}{2}\right)\sqrt{\sigma_{1\perp}^2 q_\perp^2 + \sigma_{1\parallel}^2 x_3^2}\right. \\
&\quad - \left\langle x q_\perp^2\right\rangle_{\mathrm{c}} H_1(\nu)\int \mathrm{d}q_\perp^2 \mathrm{d}x_3 \exp\left(-q_\perp^2 - \frac{x_3^2}{2}\right)\sqrt{\sigma_{1\perp}^2 q_\perp^2 + \sigma_{1\parallel}^2 x_3^2}L_1(q_\perp^2) \\
&\quad \left. + \frac{1}{2}\left\langle x x_3^2\right\rangle_{\mathrm{c}} H_1(\nu)\int \mathrm{d}q_\perp^2 \mathrm{d}x_3 \exp\left(-q_\perp^2 - \frac{x_3^2}{2}\right)\sqrt{\sigma_{1\perp}^2 q_\perp^2 + \sigma_{1\parallel}^2 x_3^2}H_2(x_3)\right]
\end{aligned}$$

and leads to the following result

$$\mathcal{N}_3^{(1)}(\nu) = \mathcal{N}_3^{(0)}(\nu)\left[\frac{1}{3!}\left\langle x^3\right\rangle H_3(\nu) + \frac{1}{2}\left(\left\langle x q_\perp^2\right\rangle + \frac{A(\beta_\sigma)/\beta_\sigma}{1-A(\beta_\sigma)}\left(\left\langle x x_3^2\right\rangle - \left\langle x q_\perp^2\right\rangle\right)\right)H_1(\nu)\right]. \tag{5.5}$$

To first order in the anisotropy parameter β_σ, we therefore get the following explicit expression

$$\mathcal{N}_3 = \frac{2e^{-\nu^2/2}\sigma_1}{\sqrt{3}\pi\sigma}\left[1 + \frac{\left\langle x^3\right\rangle}{3!}H_3(\nu) + \left(\frac{1}{3}\left\langle x q_\perp^2\right\rangle\left[1 - \frac{4}{15}\beta_\sigma\right] + \frac{1}{6}\left\langle x x_3^2\right\rangle\left[1 + \frac{8}{15}\beta_\sigma\right]\right)H_1(\nu)\right] + \mathcal{O}(\beta_\sigma^2)\,,$$

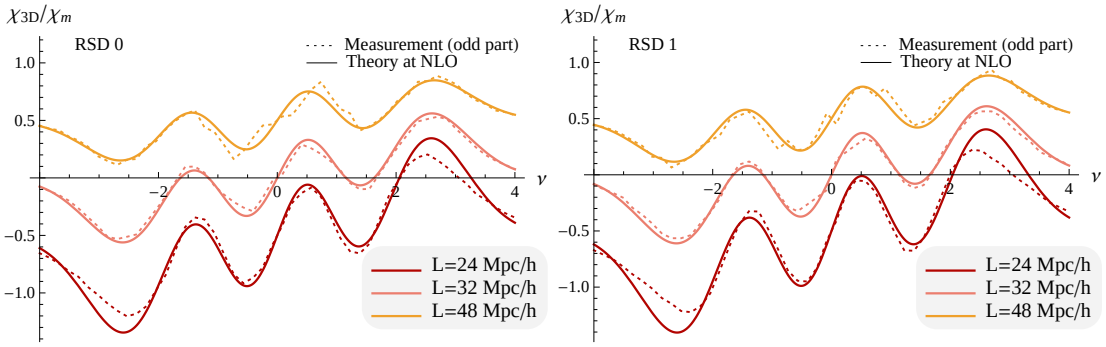


Figure 5.2: *Left panel*: first non-Gaussian correction of the 3D Euler characteristic as a function of smoothing (as labelled) measured (dotted lines) and predicted (plain line) in Horizon 4π simulation in real space. Each curve has been normalized by the maximum of the Gaussian component of the Euler characteristic. Those corresponding to different smoothing lengths have been shifted for clarity. *Right panel*: same as top left panels, but in redshift space. For these ranges of smoothing the theory predicts well the Euler characteristic to first order in non-Gaussianity, in particular for low-intermediate thresholds. The difference in the Euler characteristic introduced by redshift distortion is rather small. This figure is extracted from [Codis et al., 2013].

from which the isotropic limit of [Gay et al., 2012] is readily recovered by setting $\beta_\sigma = 0$ and $\langle xq_\perp^2 \rangle = \langle xx_3^2 \rangle = \langle xq^2 \rangle$. Anisotropy effects in $\mathcal{N}_3$ statistic are almost exclusively concentrated in the gradient terms $\propto H_1(\nu)$. This suggests, for example, that recovery, of skewness $\langle x^3 \rangle$ by fitting $H_3(\nu)$ mode to $\mathcal{N}_3$ curve will be practically unaffected by redshift distortions. In contrast, to measure anisotropic effects, one must focus on the $H_1(\nu)$ mode.

Other Minkowski functionals can be obtained in the same way as for $\mathcal{N}_3$.

Comparison to simulations The predictions appear to match the simulated result for large enough smoothing. As the smoothing length decreases, non-linearities become stronger and the prediction at first order departs from data in the tails of the distribution while for low contrast the agreement remains strong. The effect of redshift space distortions on the 3D genus seems tiny in the Horizon 4π simulation (see figure 5.2). This means that the isotropic theory is sufficient and could allow one to measure the variance of the density fluctuations using the non-Gaussian correction of topological estimators ([Gay et al., 2012]). But, it also means that there is no hope to extract more information from redshift space distortions about cosmology. However, on slices, the 2D genus *is* affected by redshift space distortions (see figure 5.3) and this effect is well taken into account by the prediction. This should allow us to measure β and therefore to possibly get more cosmological information!

From cumulants to cosmology Our aim here is to express the invariant cumulants as functions of the underlying field variance σ^2. For this purpose we need to generalize perturbation theory and its so-called geometric S_n parameters to combinations of the density field and its derivatives. The cosmic evolution of the skewness of the field is computed via integrals over configuration space of the so-called geometric Z_2 function. Here, in order to extend this result to other cumulants, we need to integrate the Z_2 function weighted by extra powers of k (that reflect the differentiations). It is known that the expectation of δ^3, scales like σ times some combinations

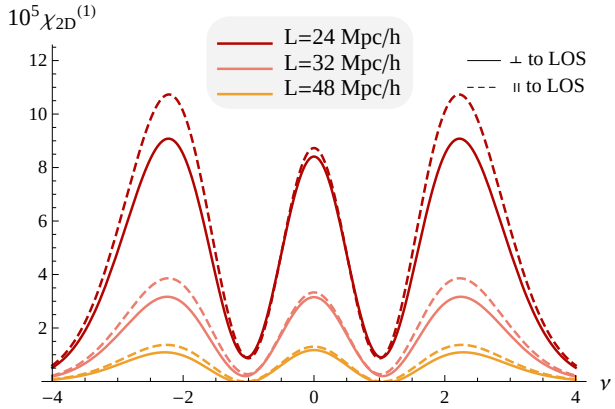


Figure 5.3: Prediction for the 2D Euler characteristic in redshift space along (dashed lines) and perpendicular (solid lines) to the line-of-sight with cumulants measured in the Horizon 4π simulation. There is a clear dependence on the angle between the slice and the line-of-sight, especially for smaller smoothing. This figure is extracted from [Codis et al., 2013].

of hypergeometric functions of the underlying power index (when Gaussian smoothings are considered). The same result holds for all moments that are cubic in the field variables. For instance, the expectation of the density field times the gradient squared becomes some other combinations of hypergeometric functions times σ. Similarly generalized quartic moments would scale like σ^2.

To be more precise, given expansion 3.32, cumulants that are cubic in the field can be computed at tree order. For conciseness, we denote simply $\int_W$ the weighted 6D integration

$$\int_W [\quad\quad] = \int \mathrm{d}^3\mathbf{k}_i \mathrm{d}^3\mathbf{k}_j [\quad\quad] W(k_i R) W(k_j R) W(|\mathbf{k}_i + \mathbf{k}_j| R)\,. \tag{5.6}$$

With this notation the cumulant of δ^3 reads for instance

$$\begin{aligned}
\left\langle \delta^3 \right\rangle &\simeq 3\left\langle (\delta^{(1)})^2 \delta^{(2)} \right\rangle = 3 D_1^4(\tau) \int_W Z_1(\mathbf{k}_1) Z_1(\mathbf{k}_2) Z_2(\mathbf{k}_3, \mathbf{k}_4) \left\langle \hat{\delta}_l(\mathbf{k}_1) \hat{\delta}_l(\mathbf{k}_2) \hat{\delta}_l(\mathbf{k}_3) \hat{\delta}_l(\mathbf{k}_4) \right\rangle , \\
&= 3(2\pi)^6 D_1^4(\tau) \left[\int_W Z_1(\mathbf{k}_1) Z_1(-\mathbf{k}_1) Z_2(\mathbf{k}_3, -\mathbf{k}_3) P(k_1) P(k_3) \right. \\
&\quad \left. + 2 \int_W Z_1(\mathbf{k}_1) Z_1(\mathbf{k}_2) Z_2(-\mathbf{k}_1, -\mathbf{k}_2) P(k_1) P(k_2) \right] ,
\end{aligned}$$

where we note that $Z_2(\mathbf{k}_3, -\mathbf{k}_3) = b_2/2$. This method can be generalized to all relevant cumulants. For instance

$$\left\langle \delta q_\perp^2 \right\rangle = (2\pi)^6 D_1^4(\tau) \left[\int_W Z_1(\mathbf{k}_1)^2 Z_2(\mathbf{k}_2, -\mathbf{k}_2) P(k_1) P(k_2) k_{1\perp}^2 \right. \tag{5.7}$$

$$\left. + 2 \int_W Z_1(\mathbf{k}_1) Z_1(\mathbf{k}_2) Z_2(\mathbf{k}_1, \mathbf{k}_2) P(k_1) P(k_2) (\mathbf{k}_{1\perp} \cdot \mathbf{k}_{2\perp} + 2k_{1\perp}^2) \right] . \tag{5.8}$$

In practice, all third order cumulants can be integrated using the decomposition of $W_G(|\mathbf{k}_1 + \mathbf{k}_2| R)$ in Legendre polynomials and Bessel functions

$$W_G(|\mathbf{k}_1 + \mathbf{k}_2| R) = \exp\left[-\frac{k_1^2 + k_2^2}{2} R^2 \right] \sum_{l=0}^{\infty} (-1)^l (2l+1) P_l \left(\frac{\mathbf{k_1} \cdot \mathbf{k_2}}{k_1 k_2} \right) I_{l+1/2}(k_1 k_2 R^2) \sqrt{\frac{\pi}{2 k_1 k_2 R^2}} ,$$

except one, given by

$$I = \int_W Z_1(\mathbf{k}_1) Z_1(\mathbf{k}_2) f\mu^2 G_2(\mathbf{k}_1, \mathbf{k}_2) P(k_1) P(k_2) \beta(\mathbf{k}_1, \mathbf{k}_2). \tag{5.9}$$

This difficulty was highlighted e.g by [Hivon et al., 1995] but was not analytically solved until [Codis et al., 2013]. For this purpose, let us use the following trick: if we introduce three different smoothing lengths $W(k_1 R_1)$, $W(k_2 R_2)$ and $W(kR)$ with a Gaussian filter, then one can see that

$$-2\frac{\partial I}{\partial R^2} = \int_W Z_1(\mathbf{k}_1) Z_1(\mathbf{k}_2) f\mu^2 k^2 G_2(\mathbf{k}_1, \mathbf{k}_2) P(k_1) P(k_2) \beta(\mathbf{k}_1, \mathbf{k}_2)\,. \tag{5.10}$$

Now the integration of equation (5.10) over $\mu^2 k^2 (= (\mathbf{k}_1 + \mathbf{k}_2)\cdot\hat{\mathbf{z}})$ is straightforward; the integration over R^2 is also straightforward and involves no constant of integration.

In the regime where standard perturbation theory holds in redshift space, all cumulants entering the expression of Minkowski functionals can thus be predicted for a given cosmological model. Measuring those functionals therefore probes the underlying cosmology by means of some particular combinations of cumulants. Redshift space analysis is in principle capable of mining more information than real space analysis. Indeed, in redshift space there is a qualitative difference between cumulants that involve the line-of-sight direction and those that involve directions orthogonal to the line-of-sight in the plane of the sky. These differences encode information about velocities, and reflect the mechanism of how these velocities originated. In principle, estimating the anisotropic part of such cumulants can be used to test the theory of gravity in the context of large-scale structures perturbation theory. 3D geometrical statistics, such as $\chi_{3\mathrm{D}}$ and $\mathcal{N}_3$ do not by themselves allow one to determine separately the line-of-sight and sky cumulants. To separate these anisotropic contributions one must analyse *slices* of 3D volume at different angles $\theta_{\mathcal{S}}$ to the line-of-sight. For instance, measuring the length of the isocontours, $\mathcal{N}_2$ (with possible cross-check from $\chi_{2\mathrm{D}}$) yields a separate handle on $\langle x q_\perp^2 \rangle$ and $\langle x x_3^2 \rangle$, while the additional analysis of the 2D Euler characteristic $\chi_{2\mathrm{D}}$, as a function of $\theta_{\mathcal{S}}$, allows us to measure $\langle J_{1\perp} q_\perp^2 \rangle$ and $\langle J_{1\perp} x_3^2 \rangle$.

Dark energy fiducial experiment As mentioned previously, the angle-dependence of the 2D Minkowski functionals allows us to probe β. Indeed, the prediction for 1 and 2D Minkowski functionals display a functional dependence on the angle $\theta_{\mathcal{S}/\mathcal{L}}$ between the 2D slice $\mathcal{S}$ or 1D line $\mathcal{L}$ and the line-of-sight. At Gaussian order first, it appears that

$$\chi_{2\mathrm{D}}^{(0)}(\nu, \theta_{\mathcal{S}}) \propto \sqrt{1 - \beta_\sigma \sin^2 \theta_{\mathcal{S}}}\,, \tag{5.11}$$

so that measuring the Gaussian part of the 2D Euler characteristic in different slices (with different orientation relative to the line-of-sight) can give access to $\beta_\sigma = 1 - \sigma_{1\perp}^2 / 2\sigma_{1\parallel}^2 = \frac{4}{5}\beta + \mathcal{O}(\beta^2)$. For example, in the most favorable case of $\theta_1 = \pi/2$ and $\theta_2 = 0$, one have direct access to

$$\frac{\chi_{2\mathrm{D}}^{(0)}(\nu, \theta_1)}{\chi_{2\mathrm{D}}^{(0)}(\nu, \theta_2)} = \sqrt{\frac{1 - \beta_\sigma \sin^2 \theta_1}{1 - \beta_\sigma \sin^2 \theta_2}} = \sqrt{1 - \beta_\sigma} = 1 - \frac{2}{5}\beta + \mathcal{O}(\beta^2)\,. \tag{5.12}$$

Contour crossing statistics have the same angle-dependence as 2D Euler characteristic at Gaussian order. The $\mathcal{N}_2$ statistics (area of isocontours) leads in turn

$$\frac{\mathcal{N}_2^{(0)}(\nu, \theta_1)}{\mathcal{N}_2^{(0)}(\nu, \theta_2)} = \frac{E(\tilde{\beta}_\sigma(\theta_1))\sqrt{1 - \tilde{\beta}_\sigma(\theta_2)}}{E(\tilde{\beta}_\sigma(\theta_2))\sqrt{1 - \tilde{\beta}_\sigma(\theta_1)}} = \frac{\pi}{2}\frac{\sqrt{1 - \beta_\sigma}}{E(\beta_\sigma)} = 1 - \frac{1}{5}\beta + \mathcal{O}(\beta^2)\,. \tag{5.13}$$

Note that beyond a simple overall amplitude effect (which is enough to measure β alone), this angle-dependence also arises in the first non-Gaussian correction. If the value of β is important on its own (e.g to study the bias or to test modifications of gravity), it is also of prime importance to measure $D(z)$ as it allows us to map the dispersion in redshift space, σ, into its value in real space, $\sigma_0 \propto D(z)$. One way to proceed is to use the Gaussian term to put constraints on β and then the non-Gaussian correction for σ (the amplitude of which is not probed by the Gaussian part). Indeed, following [Gay et al., 2012], the theory of Minkowski functionals should allow us to measure the dispersion of the field in redshift space σ from the amplitude of the departure from non-Gaussianity. As equation 5.13 demonstrates, the comparison of the 2D Minkowski functionals in planes parallel and perpendicular to the line-of-sight allows us to independently measure $\beta = f/b$. Hence Minkowski functionals in redshift space yield a geometric estimate of the real-space field dispersion, σ_0, via

$$\sigma_0 = \sigma / \sqrt{1 + 2\beta/3 + \beta^2/5}. \tag{5.14}$$

We have applied this scheme to measure β and σ_0 in a fiducial experiment. Minkowski functionals were measured from a set of 19 scale-invariant (n=-1) 256^3 dark matter simulations smoothed over 15 pixels, corresponding to $\sigma_0 = 0.18$. A first step is to use the Gaussian term of the 1D and 2D statistics, while varying the angle of the slices to constrain β. For that purpose, we extract the even part of $\mathcal{N}_2$ (to get rid of odd parity effects arising from the next-to-leading order correction) and restrict ourselves to the intermediate domain $-1.2 < \nu < 1.2$ where the Gaussian term is dominant. The resulting constraints on β are found to be $\hat{\beta} = 1.04 \pm 0.05$, and are illustrated on figure 5.4 (left panel). The same analysis on the other 1D ($\mathcal{N}_1$) and 2D statistics (2D genus) leads to similar constraints, respectively $\beta = 1.13 \pm 0.05$ and $\beta = 1.0 \pm 0.02$. The next step is to fit the first non-Gaussian correction of each statistics with PT predictions in order to constrain σ. The predictions at first order are expressed as a function of the contrast ν and σ only. The value of the free parameter σ in the model is then constrained by fitting the odd part of the data (which is dominated by the first non-Gaussian correction for intermediate contrasts). The result for $\mathcal{N}_2$ is shown in figure 5.4 (right panel) and yields $\hat{\sigma} = 0.22 \pm 0.08$. The other statistics give similar results e.g for the 3D Euler characteristic $\hat{\sigma} = 0.26 \pm 0.06$. Altogether, using σ as measured by a 3D probe, χ_{3D}, and β by a 2D statistics, $\mathcal{N}_2$, we finally get $\hat{\sigma}_0 = 0.18 \pm 0.04$, which is fully consistent with the underlying dispersion in our mocks. The accuracy on the measurement of β and $D(z)$ through σ_0 can naively be scaled to the expected accuracy for a Euclid-like survey (assuming one quarter of the sky is observed) leading to a relative 0.3% precision on β and 1.5% on $D(z)$ at redshift zero. See also [Gay et al., 2012], which translates this accuracy in terms of estimates for the dark energy parameters w_0, w_a.

Limitations and prospects In the above-mentioned dark energy fiducial experiment, it has to be noted that no account of masking, redshift evolution of S/N ratio or finite survey volume, nor comparison with other dark energy probes was attempted. One of the main limitations of the work presented in this section is that we assumed that standard perturbation theory (SPT) reproduces exactly the observations while it is known to perform somewhat poorly in redshift space. For instance, in the Horizon 4π simulation, only cumulants on scales above 50 Mpc/h were accurately predicted by SPT. Improvements are therefore critical.

Carrying out the road map sketched here while addressing the above-mentioned issues should be one of the target of the upcoming surveys that have been planned specifically to probe dark energy, either from ground-based facilities (eg BigBOSS, VST-KIDS, DES, Pan-STARRS,

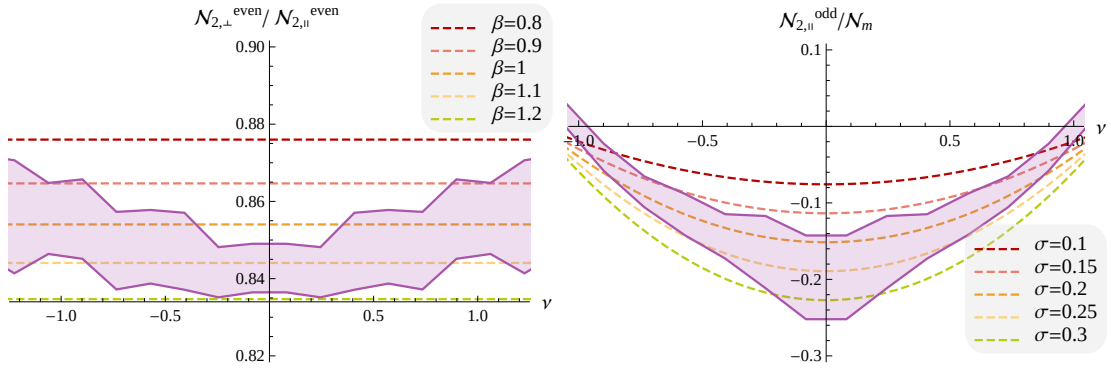


Figure 5.4: Reconstructed β from the Gaussian part of $\mathcal{N}_2$ (left panel) and σ from its non-Gaussian correction (right panel). This figure is extracted from [Codis et al., 2013].

LSST[1]) or space-based observatories (EUCLID([Laureijs et al., 2011]), SNAP and JDEM[2]).

Important improvements include

- improving perturbation theory in redshift space (while implementing variations of the streaming model (see for instance [Scoccimarro, 2004, Taruya et al., 2010]) and/or anisotropic smoothing); it would be quite interesting, for instance, to check the regime of validity of the recent model proposed by [Gil-Marín et al., 2014] in the context of topological studies;
- departing from the plane parallel approximation while constructing a full-sky prescription for non-Gaussian Minkowski functionals of realistic catalogues;
- extending the prediction to the statistics of the skeleton and walls;
- propagating to cosmic parameter estimation the residual errors;
- extending the prediction of the JPDF to N-point statistics for non-local analysis (e.g. void size and non-linear N-points peak statistics);
- exploring alternatives to the Gram-Charlier's expansion;
- deriving the statistics of *errors* on one-point statistics such as those presented in this paper;
- implementing the relevant theory on realistic mocks and demonstrating the pros and cons of geometrical probes (e.g. in the presence of masks), and contrast those to existing dark energy probes (lensing, SN1a, etc..);
- apply Minkowski functionals to the case of 2D cosmic shear maps.

5.1.1.2 The clustering of peaks

Using the same formalism as section 2.2.2.2, it is possible to go beyond the one-point statistics of peaks and study their clustering. On large scales, one can use perturbative expansions and study peak bias. On smaller scales, the two-point correlation function of peaks shows that peaks present an exclusion zone that prevent them to be infinitely close one to another and alter the power spectrum that deviates from a Poisson distribution.

[1] `http://bigboss.lbl.gov`, `http://wtww.astro-wise.org/projects/KIDS`, `https://www.darkenergysurvey.org`, `http://pan-starrs.ifa.hawaii.edu`, `http://www.lsst.org`
[2] `http://sci.esa.int/euclid`, `http://snap.lbl.gov`, `http://jdem.lbl.gov`

The exclusion region and the non-linear bias bump beyond are important for precision models of the halo-halo correlation function in the transition region between the one- and two-halo terms in the halo model. Thus, a better understanding of these regions will likely improve the modelling of the matter power spectrum or correlation function in this regime, which is very important for weak lensing or galaxy-galaxy lensing studies since the signal is large and not yet dominated by the fully non-perturbative one halo term.

The ideas and results presented in this section are the subject of an article in preparation with Tobias Baldauf, Vincent Desjacques and Christophe Pichon (see paper 5.3). Some preliminary results beyond the 1D case are also presented.

Mathematical formalism The mathematical formulation is rather simple and requires the knowledge of the joint PDF of the field and its first and second derivatives $\mathbf{X} = \{x, x_{ij}, x_i\}$ and $\mathbf{Y} = \{y, y_{ij}, y_i\}$, at two prescribed comoving locations ($\mathbf{r}_x$ and $\mathbf{r}_y$ separated by a distance $r = |\mathbf{r}_x - \mathbf{r}_y|$). This joint PDF is given by

$$\mathcal{P}(\mathbf{X}, \mathbf{Y}) = \frac{\exp\left(-\frac{1}{2}\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}^{\mathrm{T}} \cdot \mathbf{C}^{-1} \cdot \begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}\right)}{\det|\mathbf{C}|^{1/2}\,(2\pi)^{(\mathrm{D}+1)(\mathrm{D}+2)/2}}\,, \tag{5.15}$$

where $\mathbf{C}_0 \equiv \langle \mathbf{X} \cdot \mathbf{X}^{\mathrm{T}} \rangle$ and $\mathbf{C}_\gamma \equiv \langle \mathbf{X} \cdot \mathbf{Y}^{\mathrm{T}} \rangle$ are the diagonal and off-diagonal components of the covariance matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_0 & \mathbf{C}_\gamma \\ \mathbf{C}_\gamma^{\mathrm{T}} & \mathbf{C}_0 \end{bmatrix}\,. \tag{5.16}$$

All these quantities depend only on the separation vector $\mathbf{r}$ because of homogeneity. Isotropy further implies that they depend on the modulus $r = |\mathbf{r}|$ solely. From this joint PDF, the two-point correlation of (signed) critical points separated by r at threshold ν, $\xi^{\mathrm{crit}}(r, \nu)$, is given by

$$\xi^{\mathrm{crit}}(r, \nu) = \frac{\langle (n_{\mathrm{crit}}(\mathbf{r}_x, \nu) - \bar{n}_{\mathrm{crit}}(\nu))(n_{\mathrm{crit}}(\mathbf{r}_y, \nu) - \bar{n}_{\mathrm{crit}}(\nu)) \rangle}{\bar{n}^2_{\mathrm{crit}}(\nu)} \tag{5.17}$$

Therefore,

$$1 + \xi^{\mathrm{crit}}(r, \nu) = \frac{\langle \mathcal{B}_{\mathrm{crit}}(x, x_i, x_{ij}) \mathcal{B}_{\mathrm{crit}}(y, y_i, y_{ij}) \rangle}{\langle \mathcal{B}_{\mathrm{crit}}(x, x_i, x_{ij}) \rangle^2}\,, \tag{5.18}$$

where the condition for a signed critical point reads

$$\mathcal{B}_{\mathrm{crit}}(x, x_i, x_{ij}) = \det(x_{ij}) \delta_{\mathrm{D}}(x_i) \delta_{\mathrm{D}}(x - \nu)\,. \tag{5.19}$$

The denominator

$$\langle \mathcal{B}_{\mathrm{crit}}(x, x_i, x_{ij}) \rangle = \int \mathrm{d}\mathbf{X} \det(x_{ij}) \delta_{\mathrm{D}}(x_i) \delta_{\mathrm{D}}(x - \nu) \mathcal{P}(\mathbf{X})\,, \tag{5.20}$$

is the average number density of signed critical points, and

$$\begin{aligned}\langle \mathcal{B}_{\mathrm{crit}}(x, x_i, x_{ij}) \mathcal{B}_{\mathrm{crit}}(y, y_i, y_{ij}) \rangle = \iint \mathrm{d}\mathbf{X} \mathrm{d}\mathbf{Y}\, \mathcal{P}(\mathbf{X}, \mathbf{Y}) \\ \times \det(x_{ij}) \delta_{\mathrm{D}}(x_i) \delta_{\mathrm{D}}(x - \nu) \det(y_{ij}) \delta_{\mathrm{D}}(y_i) \delta_{\mathrm{D}}(y - \nu)\end{aligned} \tag{5.21}$$

is the cross-correlation. This correlation function of signed critical points differs from the correlation function of critical points because the absolute values around the determinant are dropped. The nice property of signed critical points is that since the integrant simply is a polynomial function of the variables, the integral can be fully carried out analytically. For peaks, an additional constraint on the sign of the second derivatives is required. As a consequence, the peak-peak correlation becomes

$$1 + \xi^{\rm peak}(r,\nu) = \frac{\left\langle \mathcal{B}_{\rm pk}(x,x_i,x_{ij})\mathcal{B}_{\rm pk}(y,y_i,y_{ij})\right\rangle}{\left\langle \mathcal{B}_{\rm pk}(x,x_i,x_{ij})\right\rangle^2}\,. \tag{5.22}$$

where $\mathcal{B}_{\rm pk}(x,x_i,x_{ij})$ is the peak condition given by

$$\mathcal{B}_{\rm pk}(x,x_i,x_{ij}) = |\det(x_{ij})|\delta_{\rm D}(x_i)\Theta_{\rm H}(-\lambda_i)\delta_{\rm D}(x-\nu)\,. \tag{5.23}$$

Here $|\det(x_{ij})| = -\det(x_{ij})$ because the determinant is negative at the peaks and it is understood that, for $D > 1$, $\delta_{\rm D}(x_i)$ stands for $\prod_{i\leq D}\delta_{\rm D}(x_i)$, while $\Theta_{\rm H}(-y_{ii})$ means $\prod_{l\leq D}\Theta_{\rm H}(-\lambda_l)$ with $\{\lambda_l\}_l$ the eigenvalues of the Hessian. Because of these inequalities, the integral typically is not analytical anymore.

We therefore choose to resort to Monte-Carlo methods in order to evaluate numerically equation 5.22. Namely, we draw random numbers of dimension $(\mathrm{D}+1)(\mathrm{D}+2)/4-2-2\mathrm{D}$ from the conditional probability that x_{ij} and y_{ij} satisfy the PDF, subject to the condition that $x_i = 0$ and $x = y = \nu$ (using `RandomVariate`). For each draw $^{(k)}$ if $\lambda_l(x_{ij}^{(k)}) < 0$ and $\lambda_l(y_{ij}^{(k)}) < 0$ $(l \leq D)$ we keep the sample and evaluate $\det(x_{ij}^{(k)})\det(y_{ij}^{(k)})$ and otherwise we drop it; eventually,

$$\left\langle \mathcal{B}_{\rm crit}(x,x_i,x_{ij})\mathcal{B}_{\rm crit}(x,x_i,x_{ij})\right\rangle \approx \frac{1}{N}\sum_{k\in\mathcal{S}}\det(x_{ij}^{(k)})\det(y_{ij}^{(k)})\,, \tag{5.24}$$

where N is the total number of draws, and $\mathcal{S}$ is the subset of the indexes of draws satisfying the constraints on the eigenvalues. The same procedure can be applied to evaluate the denominator, $\langle\det(x_{ij})\delta_{\rm D}(x_i)\delta_{\rm D}(x-\nu)\rangle$. Equation 5.22 then yields $\xi^{\rm peak}(r,\nu)$. This algorithm is embarrassingly parallel and can be easily generalized, for instance, to the computation of the correlation function $\xi^{\rm peak}(r,>\nu)$ of peaks *above* a given threshold in density. In practice it is fairly efficient as the draw is customized to the shape of the underlying Gaussian PDF.

1D peaks For the sake of simplicity, let us start with the theory in 1D which corresponds to a pancake geometry . The variable are $\mathbf{X} = \{x, x_1, x_{11}\}$ and $\mathbf{Y} = \{y, y_1, y_{11}\}$. Calling

$$\mathbf{C}_0 = \begin{pmatrix} 1 & 0 & -\gamma \\ 0 & 1 & 0 \\ -\gamma & 0 & 1 \end{pmatrix}\,, \quad \mathbf{C}_\gamma = \begin{pmatrix} \gamma_{00} & \gamma_{01} & \gamma_{02} \\ \gamma_{01} & \gamma_{11} & \gamma_{12} \\ \gamma_{02} & \gamma_{12} & \gamma_{22} \end{pmatrix}\,,$$

where the γ_{ij}'s represent the correlations between the field and its derivatives at two points separated by a comoving distance r, e.g. $\gamma_{22} = \langle x_{11}y_{11}\rangle$. These γ_{ij}'s are not independent. The following relations are established via integrations by part: $\gamma_{10} = -\gamma_{01}$, $\gamma_{21} = -\gamma_{12}$, $\gamma_{20} = -\gamma\gamma_{11}$.

The $\gamma_{\rm ij}(r)$ are known function of r given by the moments of the two fields and their derivatives:

$$\gamma_{ij}(r) = \frac{1}{\sigma_i\sigma_j}\int {\rm d}k\, \exp(\imath kr)(\imath k)^i(-\imath k)^j P_k(k) \quad i \leq j\,,$$

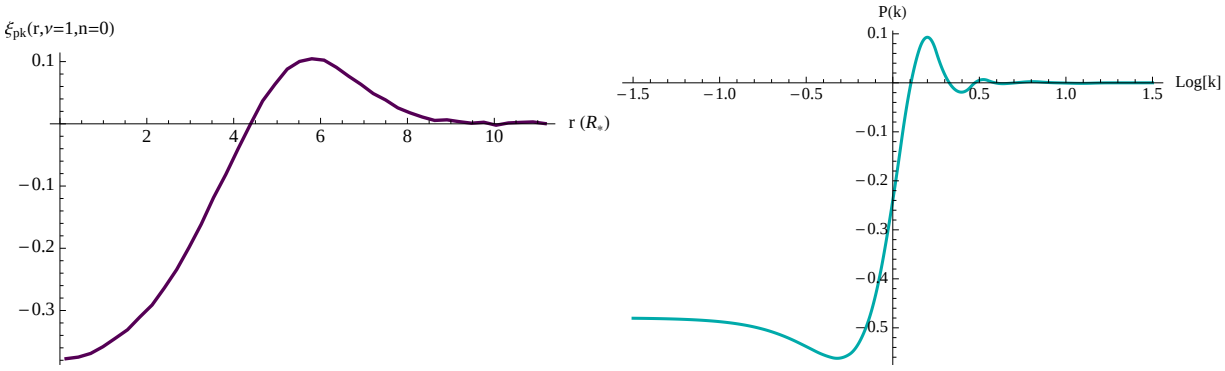


Figure 5.5: *Left panel*: 1D correlation function as a function of $r/R_\star$ for $n = 0$ at a threshold $\nu = 1$ evaluated by a Monte Carlo method. *Right panel*: Corresponding power spectrum. Because of the exclusion zone seen on the left panel, peaks are sub-Poissonian on small scales.

with $P_k(k)$ the power spectrum of δ that includes a filter function (Gaussian in this work). On expanding $\gamma_{ij}(r)$ at small separations $r \ll 1$ and substituting the spectral moments

$$\sigma_l^2 = 2\int_0^\infty k^{2l} P_k(k)\mathrm{d}k\,, \tag{5.25}$$

it follows that

$$\gamma_{ij}(r) = \frac{(-1)^{l-i}}{\sigma_i\sigma_j}\sum_{k=0}^{\infty}(-1)^k\frac{r^{2k}}{(2k)!}\sigma_{l+k}^2 \qquad (i+j=2l)$$
$$\gamma_{ij}(r) = \frac{(-1)^{1+l-i}}{\sigma_i\sigma_j}\sum_{k=0}^{\infty}(-1)^k\frac{r^{2k+1}}{(2k+1)!}\sigma_{l+k+1}^2 \quad (i+j=2l+1)$$

where r is in units of the smoothing length. The determinant of the covariance matrix is given at first order in the separation r by

$$\det\mathbf{C} = r^{18}\frac{\left(\sigma_2^6 - \left(2\sigma_1^2\sigma_3^2+\sigma_0^2\sigma_4^2\right)\sigma_2^2 + \sigma_0^2\sigma_3^4+\sigma_1^4\sigma_4^2\right)\times\left(\sigma_3^6-\left(2\sigma_2^2\sigma_4^2+\sigma_1^2\sigma_5^2\right)\sigma_3^2+\sigma_1^2\sigma_4^4+\sigma_2^4\sigma_5^2\right)}{74649600\sigma_0^4\sigma_1^4\sigma_2^4}$$
$$+\mathcal{O}(r^{19})\,.$$

Indeed, three eigenvalues of $\mathbf{C}$ are singular, respectively scaling like r^{10}, r^6 and r^2 and corresponding to the eigen-directions given by $(x-y)$, (x_1-y_1) and $(x_{11}-y_{11})$). This singularity proportional to r^{-18} is the reason why the limit $r\to 0$ is difficult to handle numerically. Analytically it means that a series expansion to eighteenth order is needed for all terms.

Integrating equation 5.18 over the six field variables leads to the two-signed critical point correlation function of same height ν and separated by r, $\xi_{1\mathrm{D}}^{\mathrm{crit}}(r,\nu)$, the result being completely analytical for any power spectrum. For the sake of readability, we do not display this result here. This function tends to zero at $\nu\to\infty$ and to minus infinity at $\nu=0$. It has two symmetric maxima at $\nu^2(r) = \sqrt{A^2+2\frac{A}{B}} - A$ where $A(r) = \frac{\gamma_{11}(1+\gamma_{11})}{2\gamma^2(1-\gamma_{11})}$ and $B(r) = \gamma_{00} + \frac{\gamma_{11}\gamma^2}{1+\gamma_{11}}$. The low-$r$ limit of this two-point correlation function of signed critical points separated by r is given by

$$1+\xi_{1\mathrm{D}}^{\mathrm{crit}}(r,\nu) = \frac{e^{\frac{\left(1-2\gamma^2\right)\nu^2}{2\left(1-\gamma^2\right)}}\left(1-\gamma_*^2\right)^{3/2}}{12\gamma\gamma_*^3\left(1-\gamma^2\right)^{5/2}}\left(1+\frac{1}{\nu^2}\times\frac{\left(1-\gamma^2\right)\left(\gamma_\#^2\gamma_*^2\gamma^2\left(2-\gamma_*^2\right)+1-\gamma_\#^2-\gamma^2\right)}{\gamma^2\gamma_\#^2\left(1-\gamma_*^2\right)^2}\right)+\mathcal{O}(r),$$

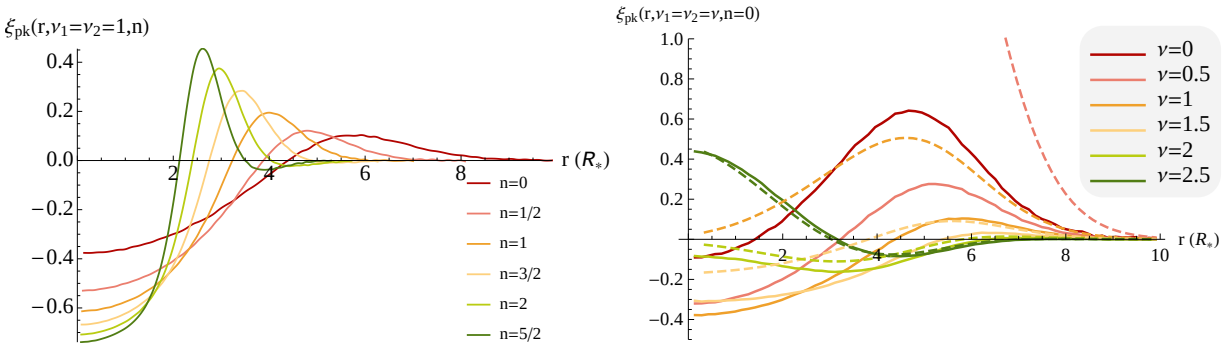


Figure 5.6: *Left panel*: 1D correlation function of peaks of height $\nu = 1$ as a function of $r/R_\star$ for different spectral index n using MCMC. *Right panel*: Same as left panel for $n = 0$ and different values of ν from 0 to 2.5 as labelled. The dashed lines correponds to the analytical 1D *signed* critical point correlation function. For $\nu > 2$, both are in perfect agreement. Note that for $\nu = 0$, the signed critical point correlation functions diverges and is not shown on this plot.

where we define the following shape parameters $\gamma = \sigma_1^2/\sigma_0/\sigma_2$, $\gamma_* = \sigma_2^2/\sigma_1/\sigma_3$ and $\gamma_\# = \sigma_3^2/\sigma_2/\sigma_4$.

For scale-invariant power spectra, $P_k(k) \propto k^n$, $n > -1$ and Gaussian smoothing, the $\gamma_{ij}(r)$ are given by

$$\begin{aligned}\gamma_{00}(r) &= {}_1F_1\left(\frac{n+1}{2};\frac{1}{2};-\frac{r^2}{4}\right)\\ \gamma_{11}(r) &= {}_1F_1\left(\frac{n+3}{2};\frac{1}{2};-\frac{r^2}{4}\right)\\ \gamma_{22}(r) &= {}_1F_1\left(\frac{n+5}{2};\frac{1}{2};-\frac{r^2}{4}\right),\end{aligned} \tag{5.26}$$

and $\gamma_{02}(r) = -\gamma\gamma_{11}(r)$. Here, r is in units of the smoothing length, ${}_1F_1$ is the Kummer confluent hypergeometric function, and

$$\sigma_l = \sqrt{\frac{\Gamma\left(l+1/2+n/2\right)}{2\pi}}. \tag{5.27}$$

The determinant of C therefore scales like $\gamma^2\left(2-\gamma^2\right)\left(1-\gamma^2\right)^{-3} r^{18}/18662400$. In the case where $n = 0$, the low-r behaviour of this function can be written as follows

$$1+\xi_{1\mathrm{D}}^{\mathrm{crit}}(r,\nu,n=0) = \frac{e^{\frac{\nu^2}{4}}\left(3\nu^2+8\right)}{8\sqrt{3}\nu^2}+\frac{e^{\frac{\nu^2}{4}}\left(128-15\nu^4\right)r^2}{1280\sqrt{3}\nu^2}+\frac{e^{\frac{\nu^2}{4}}\left(15\nu^4-64\right)r^4}{81920\sqrt{3}}+O\left(r^5\right). \tag{5.28}$$

Figure 5.5 displays on the left panel the 1D peak correlation function (evaluated by numerical integrations) as a function of $r/R_\star$ for $n = 0$ at a threshold $\nu = 1$. The result for different heights and power spectrum indices is plotted on figure 5.6. Note that as we imposed here that the peaks have the same height, the exclusion zone is reduced (not 0 at the origin). Note also that the peak-peak correlation function (solid lines) can be very accurately approximated by its genus-like counterpart (dashed lines on the right panel of figure 5.6) for $\nu > 2$. The right panel of figure 5.5 shows the corresponding power spectrum.

Figure 5.7 displays the 1D peak correlation function as a function of $r/R_\star$ for $n = 0$ for different peak heights $\nu_1 = 1 - \Delta\nu/2$ and $\nu_2 = 1 + \Delta\nu/2$, $\Delta\nu$ being between 0 and 1. It appears that

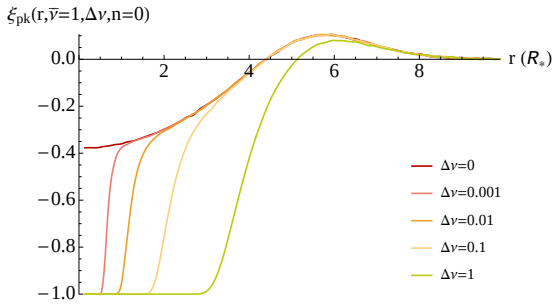


Figure 5.7: 1D correlation function as a function of $r/R_\star$ for $n = 0$ for peaks of height $\nu_1 = 1 - \Delta\nu/2$ and $\nu_2 = 1 + \Delta\nu/2$ using MCMC. The height difference varies from $\Delta\nu = 0$ to 1 as labelled.

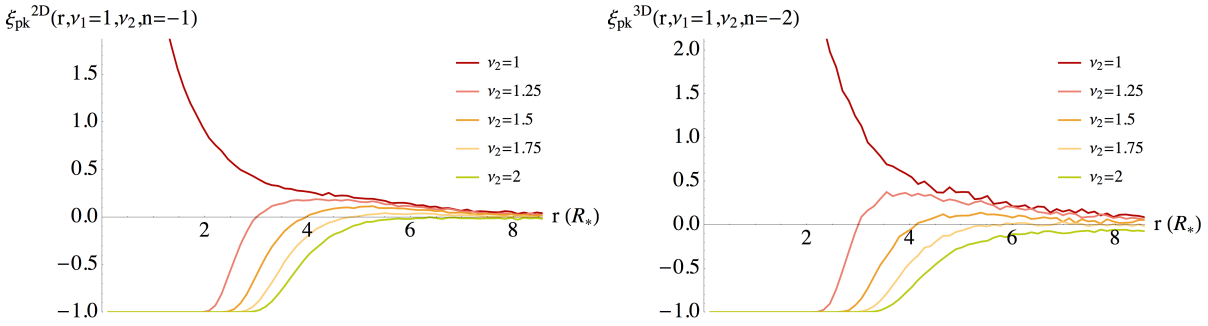


Figure 5.8: 2D (left panel) and 3D (right panel) peak correlation function as a function of r in units of $R_\star$ for $n = -1$ and -2 respectively and peak heights $\nu_1 = 1$ and ν_2 as labelled.

the peak-peak correlation function for peaks of different heights tends to 0 at small separation whereas it gives a finite non-zero limit when peaks are of the same height. The convergence of this limit thus seems to depend on the height separation. This can be easily understood as the fact that in order to have two peaks infinitely close to each other, one needs big constraints on the first and second derivatives of the field; the larger the height difference the bigger the constraints on the derivatives, the more unlikely the configuration.

The same behaviour is expected to hold for peaks on different scales (different smoothings).

Results in two and three dimensions The same formalism can be applied in two and three dimensions but the complexity of the numerical integration increases. Preliminary results are shown in figure 5.8 and 5.9 where MCMC with one million draws have been used for each point. Obviously cumulative correlation functions are more noisy compare to differential correlation function and the noise increases with the peak height as the events become rarer. Like in one dimension, one can note that the larger the height difference the more pronounced the exclusion zone.

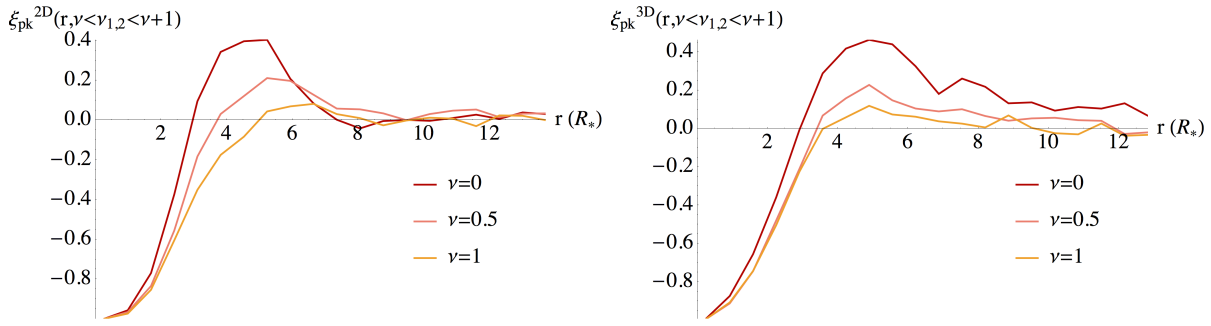


Figure 5.9: 2D (left panel) and 3D (right panel) peak correlation function as a function of r in units of the smoothing length for $n = 0$ and $-3/2$ respectively and peak height between ν and $\nu + 1$ using MCMC for different values of ν as labelled.

5.1.1.3 Multiscale peak-peak correlations in 2D

In the context of CMB analysis, it is of great interest to understand how peaks are correlated on different scales. This is particularly important for the study of extreme value statistics that can shed light on the occurence of rare events like cold spots.

In this section we therefore propose to present some preliminary results along this line that were obtained as part of an ongoing project with Stéphane Colombi. In particular, we compute the correlation function $\xi^P_{j-1,j}(\theta, \nu_{th})$ between peaks above ν_{th} of the CMB map at scales $j-1$ and j. We use the power spectrum C_ℓ as measured by the Planck collaboration filtered with the following needlet defined in harmonic space as

$$b_j(\ell) = \sqrt{h(\ell/2^j)} \tag{5.29}$$

with

$$\begin{aligned} h(x) &= 1 - s(2x), \quad x \in [1/2, 1], \\ &= s(x), \quad x \in]1, 2], \\ &= 0, \quad x \in]-\infty, 1/2[\cup]2, +\infty[, \end{aligned}$$

and

$$s(x) = 2x^3 - 9x^2 + 12x - 4\,.$$

The shape parameter at scale j is given by

$$\gamma_j = \frac{\sigma_j^2(1)}{\sigma_j(0)\sigma_j(2)} \quad \text{with} \quad \sigma_j^2(i) = \frac{1}{4\pi}\sum_\ell (2\ell+1)[\ell(\ell+1)]^i b_j(\ell)^2 C_\ell\,. \tag{5.30}$$

In the following, we use the plane-parallel approximation in the regime of small angular separations θ and we study the joint statistics of the normalized field $\nu_j = \delta_j/\sigma_j(0)$ and its normalized first $\nu_j^{,a} = \delta_j^{,a}/\sigma_j(1)$ and second derivatives $\nu_j^{,ab} = \delta_j^{,ab}/\sigma_j(2)$ collected in the 6-dimensional vector $N_j = (\nu_j, \nu_j^{,1}, \nu_j^{,11}, \nu_j^{,22}, \nu_j^{,2}, \nu_j^{,12})$. The frame used here is the one attached to the separation so that the first direction is along the separation and the second one perpendicular to it. The

auto-correlation of N_j is then trivially given by

$$\Sigma_j = \begin{pmatrix} 1 & 0 & -\frac{\gamma_j}{2} & -\frac{\gamma_j}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ -\frac{\gamma_j}{2} & 0 & \frac{3}{8} & \frac{1}{8} & 0 & 0 \\ -\frac{\gamma_j}{2} & 0 & \frac{1}{8} & \frac{3}{8} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{8} \end{pmatrix} \tag{5.31}$$

and the cross-correlation between N_{j-1} and N_j in two locations separated by an angle θ

$$\Sigma_{j-1,j} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \gamma_{1,4} & 0 & 0 \\ -\gamma_{1,2} & -\gamma_{1,3}/\gamma_j & \gamma_{2,3} & \gamma_{2,4} & 0 & 0 \\ \gamma_{1,3} & -\gamma_{2,3} & \gamma_{3,3} & \gamma_{3,4} & 0 & 0 \\ \gamma_{1,4} & -\gamma_{2,4} & \gamma_{3,4} & \gamma_{4,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\gamma_{1,4}/\gamma_j & \gamma_{2,4} \\ 0 & 0 & 0 & 0 & -\gamma_{2,4} & \gamma_{3,4} \end{pmatrix} \tag{5.32}$$

where in particular the two-point correlation function of the density field, its gradient and its Laplacian are given by

$$\xi_0(\theta) = \langle \nu_{j-1}(\vartheta)\nu_j(\vartheta+\theta)\rangle = \sum_\ell \frac{2\ell+1}{4\pi} b_{j-1}b_j C_\ell P_\ell(\cos\theta) = \gamma_{1,1}\,, \tag{5.33}$$

$$\xi_1(\theta) = \left\langle \sum_{i=1,2} \nu^i_{j-1}(\vartheta)\nu^i_j(\vartheta+\theta)\right\rangle = \sum_\ell \frac{2\ell+1}{4\pi}\ell(\ell-1) b_{j-1}b_j C_\ell P_\ell(\cos\theta) = -\frac{1}{\gamma}(\gamma_{1,3}+\gamma_{1,4})\,,$$

$$\xi_2(\theta) = \left\langle \sum_{i,i'=1,2} \nu^{ii}_{j-1}(\vartheta)\nu^{i'i'}_j(\vartheta+\theta)\right\rangle = \sum_\ell \frac{2\ell+1}{4\pi}\ell^2(\ell-1)^2 b_{j-1}b_j C_\ell P_\ell(\cos\theta) = \gamma_{3,3}+\gamma_{4,4}+2\gamma_{3,4}\,.$$

Going to a continuous description in Fourier space, where wavevectors are $\boldsymbol{\ell} = \ell(\cos\beta, \sin\beta)$ in the frame of the separation, yields for $i \leq j$

$$\gamma_{m,n}(\theta) = \frac{\int \ell\,\mathrm{d}\ell\,\mathrm{d}\beta\, C(\ell) b_{j-1}(\ell) b_j(\ell) e^{i\,l\,\theta\cos\beta} (i\,l\cos\beta)^{p_m} (i\,l\sin\beta)^{q_m} (-i\,l\cos\beta)^{p_n} (-i\,l\sin\beta)^{q_n}}{4\pi^2 \sigma_{j-1}(p_m+q_m)\sigma_j(p_n+q_n)}$$

where p_i (resp. q_i) are the number of derivatives of $N_j(i)$ along (resp. perpendicular to) the separation. At this stage, the angular integration is performed analytically and gives trivial Bessel functions and the remaining integration over the wavenumber is done numerically. Fig 5.10 shows how the small angle approximation compares to the full spherical computation of the two-point correlation function of the density field.

Consequently, one can easily compute the cross-correlation matrix $\Sigma_{j-1,j}(\theta)$ given by equation 5.32 for different values of the scale j and the angular separation θ and the resulting Gaussian probability density function of N_j is

$$P(N_j) = \frac{1}{\sqrt{(2\pi)^6 \det \Sigma_j}} \exp\left(-\frac{1}{2} N_j^t \cdot \Sigma_j \cdot N_j\right) \tag{5.34}$$

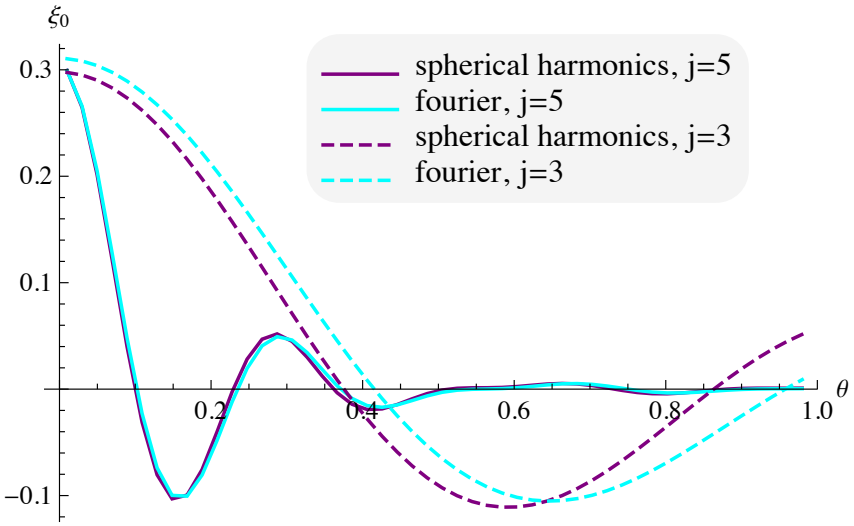


Figure 5.10: The two-point correlation function of the density field computed from equation 5.33 or Fourier transform for different values of the scale j. For large enough $j > 5$, the agreement between both methods is very good and justifies the use of Fourier space for the rest of the paper. At large scale ($j \leq 4$), this approximation breaks down.

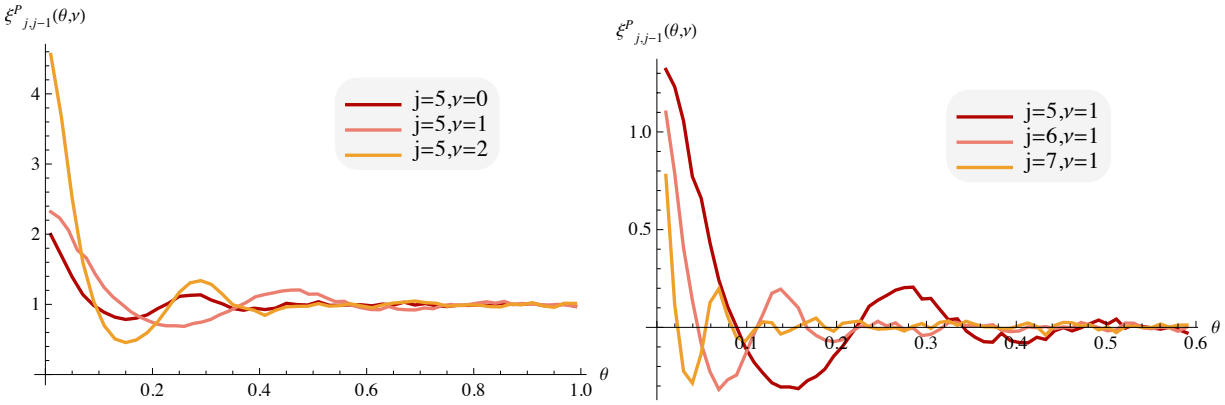


Figure 5.11: The peak-peak correlation function of the density field computed from equation 5.37 for different values of the density contrast (top panel) and different values of the scale j (bottom panel).

while the joint PDF of N_j and N'_{j-1} in two locations separated by θ becomes

$$P(N_j, N'_{j-1}) = \frac{1}{\sqrt{(2\pi)^{12}\det\Sigma}} \exp\left(-\frac{1}{2}(N_j, N'_{j-1})^t \cdot \Sigma^{-1} \cdot (N_j, N'_{j-1})\right) \tag{5.35}$$

with

$$\Sigma = \begin{pmatrix} \Sigma_j & \Sigma_{j-1,j}(\theta) \\ \Sigma^t_{j-1,j}(\theta) & \Sigma_{j-1} \end{pmatrix} . \tag{5.36}$$

The peak-peak correlation function is then given by

$$1 + \xi^P_{j-1,j}(\beta, \nu_{th}) = \frac{\langle n_{\rm pk}(\vartheta, \nu_{th}, j-1) n_{\rm pk}(\vartheta+\theta, \nu_{th}, j)\rangle_\vartheta}{\bar{n}_{\rm pk}(\nu_{th}, j)\bar{n}_{\rm pk}(\nu_{th}, j-1)} \tag{5.37}$$

where the number density of peaks above ν_{th} at scale j is given by

$$\bar{n}_{\rm pk}(\nu_{th}, j) = \int {\rm d}^6 N_j \, P(N_j) B_{\rm pk}(N_j, \nu_{th}) \tag{5.38}$$

and the joint probability of having two peaks above ν_{th} separated by θ can be written

$$\langle n_{\rm pk}(\vartheta, \nu_{th}, j-1) n_{\rm pk}(\vartheta+\theta, \nu_{th}, j)\rangle_\vartheta = \int {\rm d}^6 N_j \, {\rm d}^6 N'_{j-1} \, P(N_j, N'_{j-1}) B_{\rm pk}(N_j, \nu_{th}) B_{\rm pk}(N'_{j-1}, \nu_{th})$$

where the peak condition is written following [Bardeen et al., 1986] by adding a condition of zero gradient, of negative eigenvalues and contrast above ν_{th} while multiplying by the determinant of the Hessian matrix of the density field

$$B_{\rm pk}(N_j, \nu_{th}) = \delta_D(\nu^i_j)\Theta(\nu_j > \nu_{th}) \det \nu^{mn}_j \Theta(\det \nu^{mn}_j)\Theta(-{\rm tr}\, \nu^{mn}_j) . \tag{5.39}$$

Fig 5.11 shows the resulting two-point correlation function of peaks above ν_{th} for different values of the contrast and the scale j.

This ongoing project now requires to : i) compare the results with simulations, ii) get the correlations for a wide range of scales and peak heights and iii) possibly investigate how to compute peak-peak correlations without the plane parallel approximation.

5.1.2 Count-In-Cell statistics

A second method to accurately probe the non-linear regime of structure formation is to implement perturbation theory in a highly symmetric configuration (spherical symmetry) where non-linear solutions to the gravitational dynamical equations are known (the so-called spherical collapse model). Highly symmetric observables yield very accurate analytical predictions in the mildly non-linear regime, well beyond what is usually achievable using other estimators. Such a configuration is fulfilled when counting galaxies in concentric spheres and studying the joint distribution of these concentric galaxy densities. It is the sole known statistic which possesses this property. As such, it will be very promising to apply it to future surveys. In this section, I will show how to predict with impressive accuracy the density profiles when counting galaxies in concentric spheres in the mildly non-linear regime and investigate the effect of modified gravity models on these counts through the parametrisation of the spherical collapse.

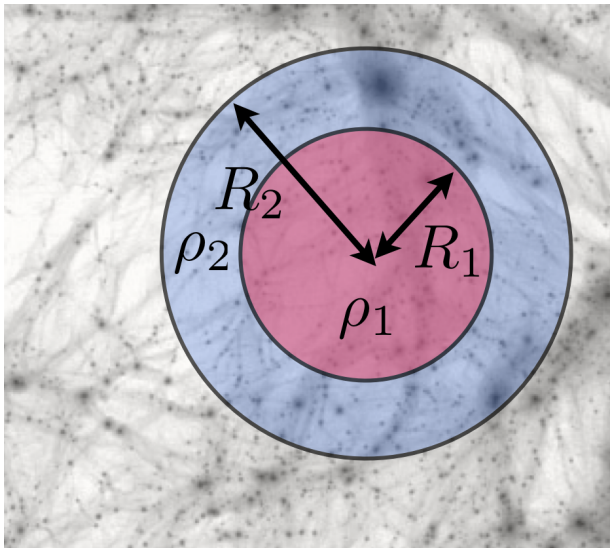


Figure 5.12: Two concentric cells of radii $R_1 < R_2$. This section aims at studying the joint statistics of the cosmic density in concentric spheres $P(\rho_1, \rho_2)$.

5.1.2.1 Mathematical formalism

The idea of count-in-cells statistics is rather simple. We take n concentric spheres, measure the density of galaxies $\hat{\rho}_i$ inside each sphere of radius R_i and study the joint statistics of those concentric densities $\mathcal{P}(\hat{\rho}_1, \dots, \hat{\rho}_n)$ (see figure 5.12). Note that here densities are rescaled so that $\langle \hat{\rho}_i \rangle = 1$. The key ingredient to predict $\mathcal{P}(\hat{\rho}_1, \dots, \hat{\rho}_n)$ from first principles is to use the moment generating function

$$\mathcal{M}_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n) = \sum_{p_1, \dots, p_n = 0}^{\infty} \langle \hat{\rho}_1^{p_1} \dots \hat{\rho}_2^{p_2} \rangle \frac{\lambda_1^{p_1} \dots \lambda_n^{p_n}}{p_1! \dots p_n!}, \tag{5.40}$$

$$= \left\langle \exp\left(\sum_i \lambda_i \hat{\rho}_i \right) \right\rangle, \tag{5.41}$$

that can be related to the *cumulant* generating function, $\varphi_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n)$, through

$$\mathcal{M}_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n) = \exp\left[\varphi_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n) \right], \tag{5.42}$$

so that

$$\exp\left[\varphi_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n) \right] = \int \mathrm{d}\hat{\rho}_1 \dots \mathrm{d}\hat{\rho}_n \, \mathcal{P}(\hat{\rho}_1, \dots, \hat{\rho}_n) \exp\left(\sum_i \lambda_i \hat{\rho}_i \right). \tag{5.43}$$

From equation 5.43, the joint PDF of concentric densities can easily be recovered by inverse Laplace transform of the cumulant generating function

$$\mathcal{P}(\hat{\rho}_1, \dots, \hat{\rho}_n) = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda_1}{2\pi \mathrm{i}} \dots \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda_n}{2\pi \mathrm{i}} \exp\left(- \sum_i \hat{\rho}_i \lambda_i + \varphi_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n) \right). \tag{5.44}$$

Cumulant generating function In order to predict the cumulant generating function, one is led to consider its *Legendre transform* in the quasi-linear regime

$$\Psi_{R_1 \dots R_n}(\rho_1, \dots, \rho_n) = \sum_i \lambda_i \rho_i - \varphi_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n), \tag{5.45}$$

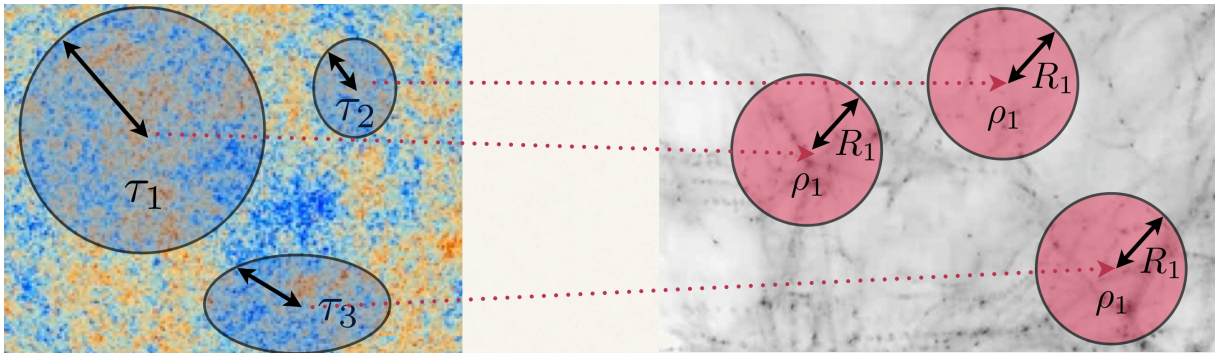


Figure 5.13: A same present-time density (right-hand side) can arise from very different initial configurations (left-hand side).

where the densities ρ_i are determined implicitly by the stationarity conditions

$$\lambda_i = \frac{\partial}{\partial \rho_i} \Psi_{R_1 \dots R_n}(\rho_1, \dots, \rho_n)\,, \quad i = 1, \dots, n\,. \tag{5.46}$$

In the limit of zero variance, the Legendre transforms taken at two different times, $\Psi(\rho_1, \dots, \rho_n; \eta)$ and $\Psi'(\rho_1, \dots, \rho_n; \eta')$, take the *same* value

$$\Psi_{R_1 \dots R_n}(\rho_1, \dots, \rho_n; \eta) = \Psi_{R'_1 \dots R'_n}(\rho'_1, \dots, \rho'_n; \eta')\,, \tag{5.47}$$

provided that the mass is preserved $\rho_i R_i^3 = \rho'_i R_i'^3$, and that ρ'_i and ρ_i are linked together through the nonlinear dynamics of the spherical collapse.

Indeed, it is always possible to express any ensemble average in terms of the statistical properties of the initial density field so that we can formally write

$$\exp\left[\varphi\right] = \int \mathcal{D}\tau_1 \dots \mathcal{D}\tau_n\, \mathcal{P}(\tau_1, \dots, \tau_n) \exp\left(\sum_i \lambda_i \rho_i(\tau_i)\right)\,. \tag{5.48}$$

As the present-time densities ρ_i can arise from different initial contrasts (see figure 5.18), the above-written integration is therefore a path integral (over all the possible paths from initial conditions to present-time configuration) with measure $\mathcal{D}\tau_1 \dots \mathcal{D}\tau_n$ and known initial statistics $\mathcal{P}(\tau_1, \dots, \tau_n)$. Let us assume here that the initial PDF is Gaussian so that,

$$\mathcal{P}(\tau_1, \dots, \tau_n) \mathrm{d}\tau_1 \dots \mathrm{d}\tau_n = \frac{\sqrt{\det \Xi}\exp\left[-\Psi(\tau_1, \dots, \tau_n)\right]}{2\pi} \mathrm{d}\tau_1 \dots \mathrm{d}\tau_n\,, \tag{5.49}$$

with Ψ then a quadratic form.

In the regime where the variance of the density field is small, equation 5.48 is dominated by the path corresponding to the most likely configurations. As the configuration is spherically symmetric, this most likely path should also respect spherical symmetry. It is therefore bound to obey the spherical collapse dynamics. Within this regime equation 5.48 becomes

$$\exp\left[\varphi\right] \simeq \int \mathrm{d}\tau_1 \dots \mathrm{d}\tau_n\, \mathcal{P}(\tau_1, \dots, \tau_n) \exp\left(\sum_i \lambda_i \zeta_{\mathrm{SC}}(\tau_i)\right)\,, \tag{5.50}$$

where the most likely path, $\rho_i = \zeta_{\mathrm{SC}}(\eta, \tau_i)$ is the one-to-one spherical collapse mapping between one final density at time η and one initial density contrast. The integration on the right-hand

side of equation 5.50 can now be carried by using a steepest descent method, approximating the integral as its most likely value, where $\sum_i \lambda_i \rho_i(\tau_i) - \Psi(\tau_1, \ldots, \tau_n)$ is stationary. It eventually leads to the fundamental relation (5.47) which, when applied to an arbitrarily early time η', yields a relation between $\Psi(\rho_1, \ldots, \rho_n; \eta)$ and the statistical properties of the *initial* density fluctuations. For Gaussian initial conditions, $\Psi(\rho_1, \ldots, \rho_n; \eta_i)$ can easily be calculated and expressed in terms of elements of covariance matrices,

$$\Psi_{R_1 \ldots R_n}(\rho_1, \ldots, \rho_n; \eta_i) = \frac{1}{2} \sum_{i,j \leq 2} \Xi_{ij} (\rho_i - 1)(\rho_j - 1) \,, \tag{5.51}$$

where Ξ_{ij} is the *inverse* of the matrix of covariances, $\Sigma_{ij} = \langle \tau_i \tau_j \rangle$, between the initial density contrasts $\tau_i = \rho_i - 1$ in the n concentric spheres of radii R_i. One can then write the cumulant generating function at any time through the spherical collapse mapping between one final density at time η in a sphere of radius R_i and one initial contrast in a sphere centered on the same point and with radius $R_i' = R_i \rho_i^{1/3}$ (so as to encompass the same total mass); it can be written formally as

$$\rho_i = \zeta_{\rm SC}(\eta, \tau_i) \approx \frac{1}{(1 - D_+(\eta)\tau/\nu)^\nu} \,, \tag{5.52}$$

where, for the sake of simplicity, we use here a simple prescription, with $D_+(\eta)$ the linear growth factor and $\nu = 21/13$ to reproduce the high-z skewness.

The joint PDF In principle, the formalism described here allows one to predict the PDF of concentric densities in the quasi-linear regime. Recall that only $\Psi_{R_1,\ldots,R_n}(\rho_1, \ldots, \rho_n)$ is easily computed. The statistically relevant cumulant generating function, $\varphi_{R_1 \ldots R_n}(\lambda_1, \ldots, \lambda_n)$, is only accessible via equation 5.45 through an inverse Legendre transform which brings its own complications. In particular note that all values of λ_i are not accessible due to the fact that the ρ_i – λ_i relation cannot always be inverted. The PDF is then accessible through an inverse Laplace transform which can be computed analytically in some regimes or from a numerical point of view. However the inverse Laplace transform generically relies on the analytic continuation of the predicted cumulant generating function in the complex plane and therefore requires a good knowledge of the analytic properties of $\varphi(\lambda)$. Its numerical implementation also appears extremely challenging and is increasingly complex with the number of cells considered. It is only recently that the two-cell PDF was successfully computed (see paper 5.5). However, an analytical low-density approximation can be found in the general n-cell case. It is based on the use of the saddle point approximation of equation 5.44 assuming the overall variance is small. It leads to the following condition that should be met at the saddle point $\{\lambda_s\}_i$

$$\frac{\partial}{\partial \lambda_k} \left[\sum_i \lambda_i \hat{\rho}_i - \varphi(\{\lambda_i\}) \right] = 0 \,, \tag{5.53}$$

which leads to

$$\hat{\rho}_i = \rho_i(\{\lambda_k\}) \,, \tag{5.54}$$

and with the constraint that

$$\det \left[\frac{\partial^2 \Psi}{\partial \rho_k \partial \rho_l} \right] > 0 \tag{5.55}$$

at the saddle point position. The resulting expression for the density PDF reads

$$\mathcal{P}(\{\hat{\rho}_k\}) = \frac{1}{(2\pi)^{n/2}} \sqrt{\det \left[\frac{\partial^2 \Psi}{\partial \hat{\rho}_k \partial \hat{\rho}_l} \right]} \exp\left[-\Psi(\{\hat{\rho}_k\}) \right] . \tag{5.56}$$

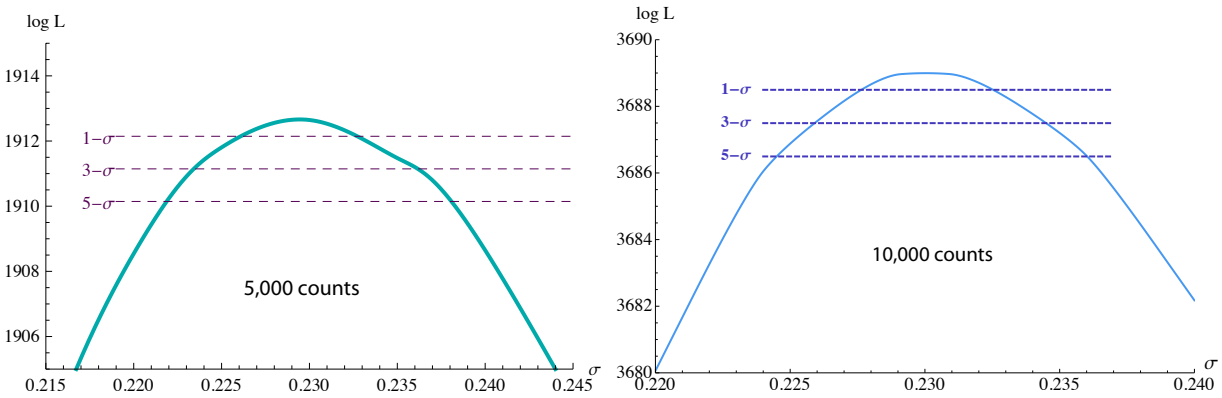


Figure 5.14: Log-likelihood for a fiducial experiment involving 5,000 (left panel) and 10,000 (right panel) concentric spheres of 10 and 11 Mpc/h measured in our simulation. The model here only depends on the variance σ (ν – that parametrizes the spherical collapse – and n – the spectral index – are fixed). The contours at 1, 3 and 5 sigmas centered on the true value 0.23 are displayed with dark blue dashed lines. The same experiment can be carried out when the three parameters vary.

This analytic expression is expected to be an approximate form for the exact PDF in underdense regions. Approximate forms in the large-density regime can also be found as shown in the coming sections for one and two cells.

Eventually, this PDF of concentric densities is sensitive to cosmology through two ingredients: the spherical collapse dynamics and the linear power-spectrum, $P_k^{\rm lin}$, (via the covariance matrix, $\Sigma_{ij} = \int P_k^{\rm lin}(k) W(R_i k) W(R_j k) d^3k/(2\pi)^3$). It can thus serve as a statistical indicator to test gravity and dark energy models and/or probe key cosmological parameters. As an illustration, I implemented a fiducial dark energy experiment on counts derived from ΛCDM simulations as seen on Fig 5.14. With 10,000 sets of two concentric spheres of 10 and 11 Mpc/h, the experiment mimics the precision expected from a survey of useful volume of about $(350h^{-1}{\rm Mpc})^3$ which is found to be at the percent level.

We now propose to give more details about the case of one and two cells in the following sections 5.1.2.2 and 5.1.2.3. Note that in addition to the cosmic density statistics, the n-cell formalism presented here also allows to study the slope of the density field defined as the difference of density between two adjacent cells.

5.1.2.2 One-cell density

For illustrative purpose, let us focus first on the one-cell formalism (e.g. [Bernardeau, 1994b]). In the one-cell case,

$$\Psi(\rho) \equiv \frac{1}{2\sigma^2(R\rho^{1/3})}\tau(\rho)^2\,, \tag{5.57}$$

where

$$\sigma^2(r) = \langle \tau(<r)\tau(<r)\rangle. \tag{5.58}$$

The Legendre transform is then straightforward and $\varphi(\lambda)$ takes the form

$$\varphi(\lambda) = \lambda\rho - \frac{1}{2\sigma^2(R\rho^{1/3})}\tau(\rho)^2\,, \tag{5.59}$$

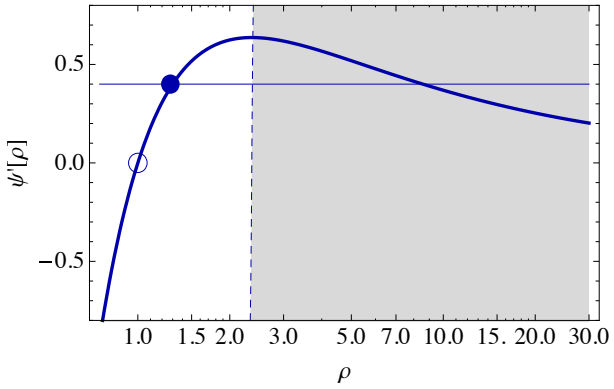


Figure 5.15: A graphical representation of the 1D stationary condition $\lambda = \Psi'[\rho]$. There is a maximum value for λ that corresponds to a critical value ρ_c for ρ defined in equation 5.62. This figure is from [Bernardeau et al., 2014].

with ρ computed implicitly as a function of λ via

$$\lambda = \frac{\partial}{\partial \rho}\Psi(\rho). \tag{5.60}$$

Eventually, the density PDF reads

$$\mathcal{P}(\hat{\rho}) = \int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} \frac{\mathrm{d}\lambda}{2\pi\mathrm{i}} \exp(-\lambda\hat{\rho} + \varphi(\lambda))\,, \tag{5.61}$$

which can be evaluated either numerically or analytically in the low (see equation 5.56) and high-density regime. We propose now to present this high-density approximation of the density PDF.

Figure 5.15 shows graphically that there is a maximum value for λ, λ_c, that can be reached, so that the Legendre Transform of Ψ is not defined for $\lambda > \lambda_c$. It corresponds to a value $\rho = \rho_c$. At this location we have

$$0 = \Psi''[\rho_c]\,, \qquad \lambda_c = \Psi'[\rho_c]\,. \tag{5.62}$$

Note that at $\rho = \rho_c$, Ψ is regular (in particular, the corresponding singular behavior in $\varphi(\lambda)$ is not related to any singularity of the spherical collapse dynamics). The function $\varphi(\lambda)$ can thus be expanded at this point. First, equation 5.60

$$\lambda = \lambda_c + \sum_i \frac{(\rho - \rho_c)^i}{i!}\pi_{i+1} \tag{5.63}$$

can be inverted as a series near (ρ_c, λ_c) (where equation 5.62 holds). For instance, at order $i = 5$ and defining $\Delta\lambda = \lambda - \lambda_c$, it gives

$$\rho = \rho_c + \sqrt{\frac{2}{\pi_3}}\sqrt{\Delta\lambda} - \frac{\pi_4\Delta\lambda}{3\pi_3^2} + \frac{\left(5\pi_4^2 - 3\pi_3\pi_5\right)\Delta\lambda^{3/2}}{18\sqrt{2}\pi_3^{7/2}} + \frac{\left(45\pi_3\pi_5\pi_4 - 9\pi_3^2\pi_6 - 40\pi_4^3\right)\Delta\lambda^2}{270\pi_3^5} \tag{5.64}$$

so that $\varphi(\lambda)$ can finally be expanded around λ_c by plugging equation 5.64 into

$$\varphi(\lambda) = \lambda\rho - \Psi(\rho). \tag{5.65}$$

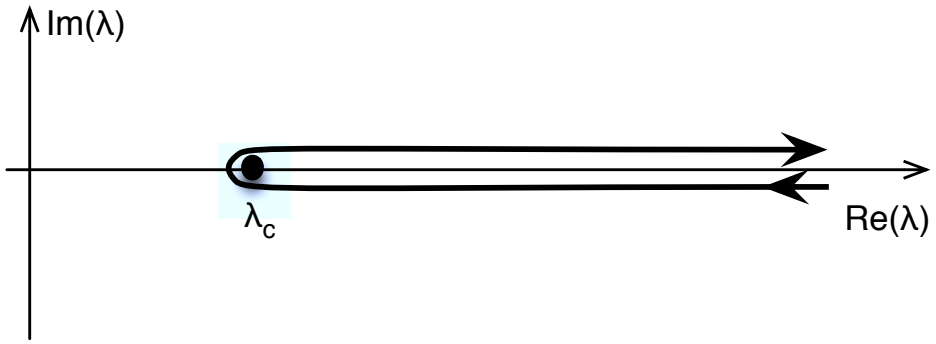


Figure 5.16: The path line in the λ complex plane for the computation of the large density asymptotic forms.

We give here a whole set of sub-leading terms that we will take advantage of in the following,

$$\begin{aligned}\varphi(\lambda) &= \varphi_c + \Delta\lambda\rho_c + \frac{2}{3}\sqrt{\frac{2}{\pi_3}}\Delta\lambda^{\frac{3}{2}} - \frac{\pi_4\Delta\lambda^2}{6\pi_3^2} + \frac{\left(5\pi_4^2 - 3\pi_3\pi_5\right)\Delta\lambda^{\frac{5}{2}}}{45\sqrt{2}\pi_3^{7/2}} \\ &- \frac{\left(40\pi_4^3 - 45\pi_3\pi_5\pi_4 + 9\pi_3^2\pi_6\right)\Delta\lambda^3}{810\pi_3^5} \\ &+ \frac{\left(385\pi_4^4 - 630\pi_3\pi_5\pi_4^2 + 168\pi_3^2\pi_6\pi_4 + 3\pi_3^2\left(35\pi_5^2 - 8\pi_3\pi_7\right)\right)\Delta\lambda^{\frac{7}{2}}}{7560\sqrt{2}\pi_3^{13/2}} + \dots \end{aligned} \tag{5.66}$$

where $\pi_n = \partial^n\Psi/\partial\rho^n(\rho_c)$. It is to be noted that the leading singular term scales like $(\lambda - \lambda_c)^{3/2}$. The coefficients π_i are all related to the function Ψ and are therefore (cosmological) model dependent. The consequence for the density PDF is as follows. When the condition $\hat{\rho} < \rho_c$ is not satisfied, the singular behavior of φ near λ_c dominates the integral in the complex plane. This leads to the following expression for $\mathcal{P}(\hat{\rho})$

$$\mathcal{P}(\hat{\rho}) \approx \exp\left(\varphi_c - \lambda_c\hat{\rho}\right)\left(\frac{3\,\Im(a_{\frac{3}{2}})}{4\sqrt{\pi}\left(\hat{\rho}-\rho_c\right)^{5/2}} + \frac{15\,\Im(a_{\frac{5}{2}})}{8\sqrt{\pi}\left(\hat{\rho}-\rho_c\right)^{7/2}} + \frac{105\left(\Im(a_{\frac{3}{2}})a_2 + \Im(a_{\frac{7}{2}})\right)}{16\sqrt{\pi}\left(\hat{\rho}-\rho_c\right)^{9/2}} + \dots\right), \tag{5.67}$$

where a_j are the coefficients in front of $(\lambda - \lambda_c)^j$ in equation 5.66. Indeed, the derivation of the rare event tail of the density PDF for large positive densities is based on the inverse Laplace transform of the generating function $\varphi(\lambda)$ when it is dominated by its singular part, i.e. for $\lambda \approx \lambda_c$. In this case the complex plane contour is pushed along the real axis wrapping around the singular value λ_c as depicted on figure 5.16. The general form for the density PDF is expressed using the form (5.66) following the path shown on figure 5.16. As the contributions from the two branches of the path lines are complex conjugate, it eventually leads to the form,

$$\begin{aligned}P(\hat{\rho}) &\approx \Im\left\{\int_{\mathrm{i}\epsilon+\lambda_c}^{\mathrm{i}\epsilon+\infty}\frac{\mathrm{d}\lambda}{\pi}\exp[\varphi_c - \lambda_c\hat{\rho} - (\lambda-\lambda_c)(\hat{\rho}-\rho_c)]\right. \\ &\left.\times\left[1 + a_{3/2}(\lambda-\lambda_c)^{3/2} + \dots\right]\right\}, \end{aligned} \tag{5.68}$$

where we keep only the dominant singular part in $\varphi(\lambda)$ and where $\Im$ denotes the imaginary part. This integral can easily be computed and it leads to,

$$P(\hat{\rho}) \approx \exp\left(\varphi_c - \lambda_c\hat{\rho}\right)\left(\frac{3\,\Im(a_{\frac{3}{2}})}{4\sqrt{\pi}\left(\hat{\rho}-\rho_c\right)^{5/2}} + \dots\right). \tag{5.69}$$

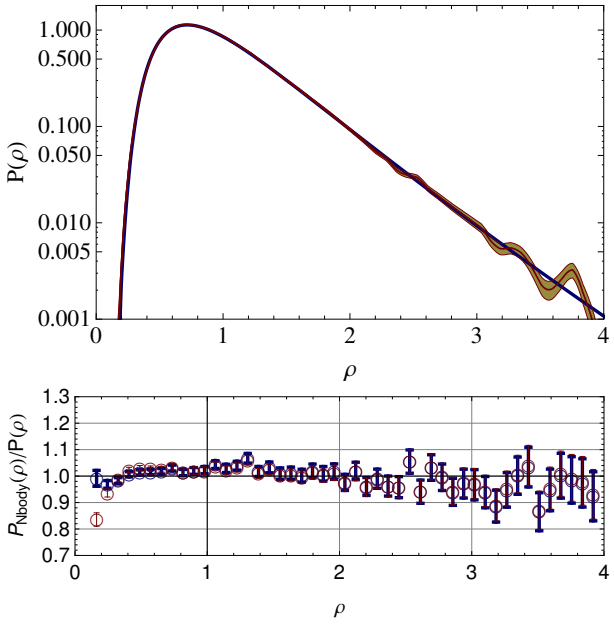


Figure 5.17: Comparison with simulations (top) with residuals (bottom). The solid line is the theoretical prediction computed for a variance of $\sigma_R^2 = 0.47$ as measured in the simulation, a power law index of $n = -1.576$ and a running parameter $\alpha = 0.439$ corresponding to the input linear power spectrum. The measured PDF in the simulation is shown as a band corresponding to its 1–σ error bar (but different data points are correlated). The residuals show the ratio of measured PDF in bins with the predictions (computed in bins as well). This figure is from [Bernardeau et al., 2014].

Sub-leading contributions can be computed in a similar way when $\exp(\varphi(\lambda))$ is expanded to higher order. Note that by symmetry, only half integer terms that appear in this expansion will actually contribute to the final density PDF given by equation 5.67. This PDF has an exponential cut-off at large $\hat{\rho}$ scaling like $\exp(\lambda_c\hat{\rho})$. It also gives a direct transcription of why $\varphi(\lambda)$ becomes singular: for values of λ that are larger than λ_c, the integral $\int \mathrm{d}\hat{\rho}\,\mathcal{P}(\hat{\rho})\exp(\lambda\hat{\rho})$ is not converging.

However, the asymptotic forms at low and high denisties are not accurate for the full range of density values; in general one has to rely on numerical integrations in the complex plane which can be done accurately and quickly in the one-cell case. In figure 5.17, we explicitly show the comparison between our predictions and a simulation. The predictions show a remarkable agreement with the measured PDF! In particular the predictions reproduce with an extremely good accuracy the PDF tails in both the low density and high density regions. The plot of the residuals shows that the predictions are at the percent level over a large range of density values. Up to $\sigma = 0.64$ ($\sigma^2 = 0.41$), we see no significant departure from the results of the simulation in the whole range of available densities, that is in particular up to about the 5σ rare event in the high density tail. This success is to be contrasted with the Edgeworth expansion approach which breaks for $|\delta| \geq \sigma$.

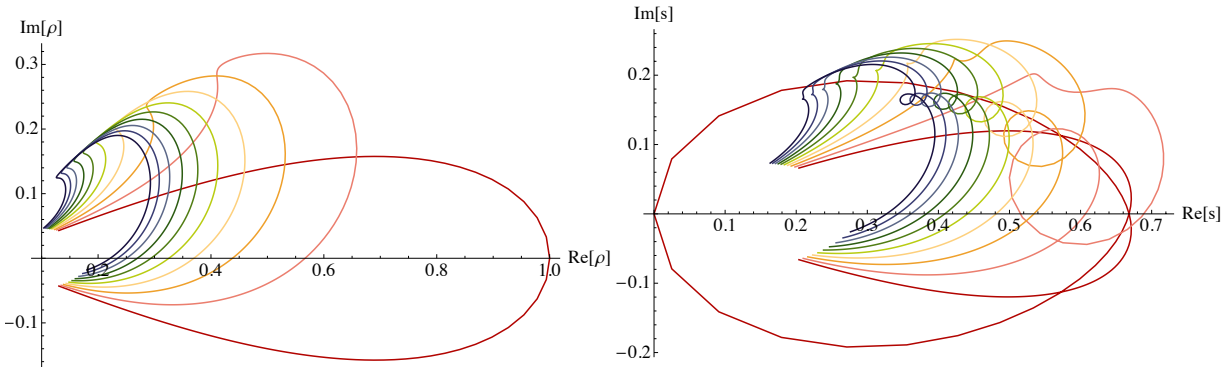


Figure 5.18: Contour path of the density ρ (top panel) and the slope s (bottom panel) when describing the imaginary path of λ at fixed μ (different colours) and conversely when describing the imaginary path of μ at fixed λ (same colours) .

5.1.2.3 Two-cell density and slope

Let us now turn to the two-cell case in which the PDF is recovered via the following inverse Laplace transform

$$\mathcal{P} = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda_1}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda_2}{2\pi\mathrm{i}} \exp(-\sum_{i=1,2} \hat{\rho}_i\lambda_i + \varphi(\lambda_1,\lambda_2)) \,. \tag{5.70}$$

From equation 5.70, it is straightforward to deduce the joint PDF, $\hat{\mathcal{P}}(\hat{\rho},\hat{s})$, for the density, $\hat{\rho}=\hat{\rho}_1$ and the slope $\hat{s} \equiv (\hat{\rho}_2 - \hat{\rho}_1)R_1/\Delta R$, ΔR being $R_2 - R_1$, as

$$\hat{\mathcal{P}}(\hat{\rho},\hat{s}) = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\mu}{2\pi\mathrm{i}} \exp(-\hat{\rho}\lambda - \hat{s}\mu + \varphi(\lambda,\mu)) \,, \tag{5.71}$$

with $\lambda = \lambda_1 + \lambda_2$, $\mu = \lambda_2 \Delta R/R_1$. Following this definition, $\varphi(\lambda,\mu)$ is also the Legendre transform of $\Psi(\hat{\rho}_1, \hat{s} = (\hat{\rho}_2 - \hat{\rho}_1)\, R_1/\Delta R)$.

In order to numerically compute equation 5.70 and therefore have the knowledge of the full two-cell PDF, we simply choose the imaginary path $(\lambda_1,\lambda_2) = \mathrm{i}(n_1\Delta\lambda, n_2\Delta\lambda)$ where n_1 and n_2 are (positive or negative) integers and the step $\Delta\lambda$ has been set to 0.15. The maximum value of λ_i used here is 75 resulting into a discretisation of the integrand on 1000^2 points. The resulting contour path of the density and the slope is shown on figure 5.18. I have also investigated the convergence of our numerical scheme by varying the number of points and showed that the proposed numerical integration of the slope PDF has reached 1% precision.

Figure 5.19 compares the result of the numerical integration of equation 5.70 to dark matter simulations. Once the two-cell PDF is known, conditional statistics can also be easily computed. This formalism allows to weigh non-uniformly different regions of the universe making possible to take into account the fact that the noise structure in surveys is not homogenous. For instance, low density regions are probed by fewer galaxies. Conversely, on dynamical grounds, we also expect the level of non-linearity in the field to be in-homogenous: low density regions are less non-linear. Hence it is of interest to build statistical estimators which probe the mildly non linear regime and that can be tuned to probe subsets of the field, offering the best compromise between these constraints. In particular, in paper 5.5, we computed density profiles restricted to underdense and overdense regions at various redshifts. Figure 5.20 for instance shows those

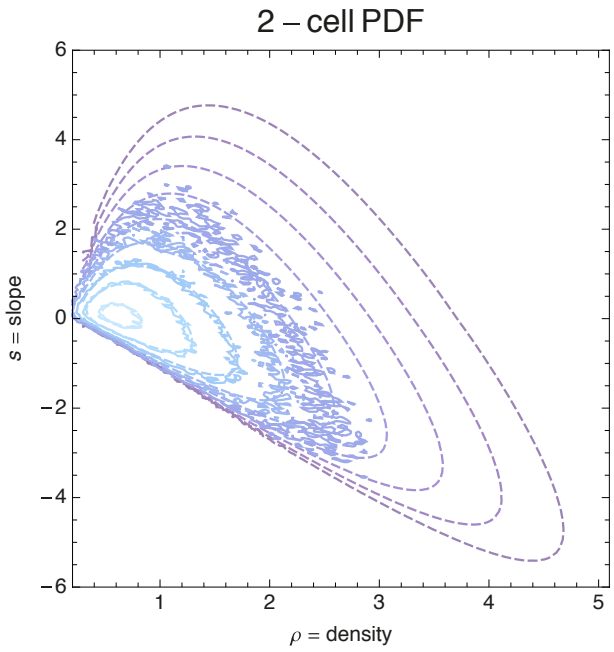


Figure 5.19: Joint PDF of the slope (s) and the density (ρ) as given by equation 5.71 for two concentric spheres of radii $R_1 = 10$ Mpc/h and $R_2 = 11$ Mpc/h at redshift $z = 0.97$. Dashed contours corresponds to Log $\mathcal{P} = 0, -1/2, -1, \cdots -3$ for the theory. The corresponding measurements are shown as a solid line. This plot is from [Bernardeau et al., 2015].

profiles at redshift $z = 0.65$. A very good agreement with simulations is found with some slight departures in the large slope tail of the distribution. As expected, the underdense slope PDF peaks towards positive slope, while the overdense PDF peaks towards negative slope. The constrained negative tails are more sensitive to the underlying constraint, providing improved leverage for measuring the underlying cosmological parameters.

5.1.2.4 Prospects

In short, count-in-cells statistics prove to be a promising cosmological probe to implement on future large surveys like Euclid. The agreement between predictions and simulations are shown to be very good, including in the quasi-linear regime where standard perturbation theory normally fails. One open question would be to estimate how many concentric cells should be used to get an optimal constraint for a given set of cosmic parameters but the answer to this question will probably depend on the geometry of the available survey. However, it has to be noted that an effective implementation of 3D count-in-cell as a cosmological test would be far fetched. In particular galaxy catalogues in z-space break the local spherical symmetry in a complex way making the application of such method impractical. One way to avoid this problem is to stick to observations for which this method is applicable, such as *projected* densities along the line of sight. It can be done either in the context of cosmic shear observations or for photometric like redshift surveys. In both cases the point is not to reconstruct the spherical 3D statistics but the circular 2D statistics for which the whole method should be applicable following early investigations. The accuracy of the predictions have still to be assessed in this context and will be the subject of future works. Another missing piece that can be incorporated is the large distance correlation of statistical indicators such as profiles and constrained profiles. Following

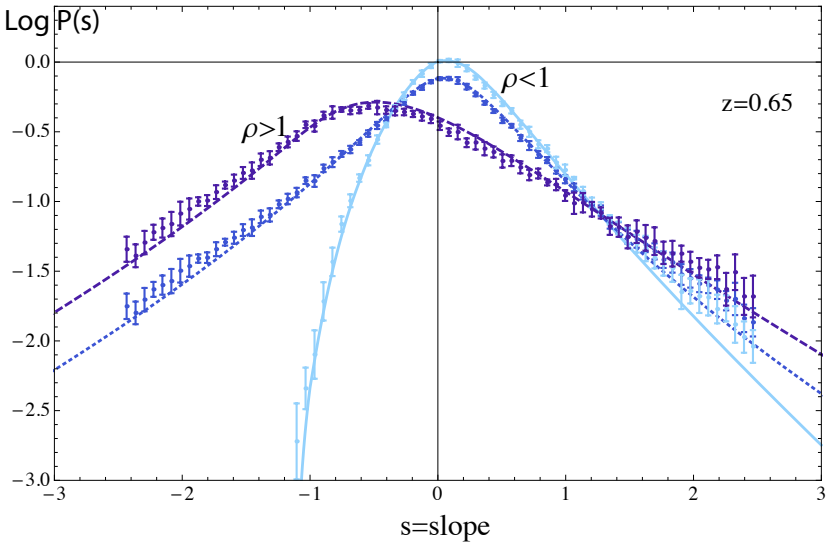


Figure 5.20: Density profiles in underdense (solid light blue), overdense (dashed purple) and all regions (dashed blue) for cells of radii $R_1 = 10$ Mpc/h and $R_2 = 11$ Mpc/h at redshift $z = 0.65$. Predictions are successfully compared to measurements in simulations (points with error bars).

[Bernardeau, 1995], it is indeed within reach of this formalism to compute such quantities. We would then have a fully working theory that could be exploited in real data sets.

5.1.3 Conclusion

In the context of high-precision cosmology, it is worth developing new observables that i) can be predicted from first principles; ii) are robust (noise, bias,...); iii) can be computed in the mildly non-linear regime of structure formation. We have shown that promising such observables include topological and geometrical estimators (Minkowski functionals, peak statistics) and count-in-cells statistics (density PDF, void profiles,...).

All the ideas presented in this section are extensively described in four (three published and one in preparation) papers I have co-signed and that are copied below. Paper 5.2 do the theory of Minkowski functionals for weakly non-Gaussian fields in redshift space, paper 5.3 then studies the exclusion zone around peaks. Count-in-cell statistics is finally the subject of papers 5.4 and 5.5 which describe the statistics of cosmic densities and slopes.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **435,** 531–564 (2013) doi:10.1093/mnras/stt1316
Advance Access publication 2013 August 14

Non-Gaussian Minkowski functionals and extrema counts in redshift space

S. Codis,[1,2]⋆ C. Pichon,[1,2] D. Pogosyan,[3] F. Bernardeau[1,2] and T. Matsubara[4]

[1]*Institut d'Astrophysique de Paris & UPMC (UMR 7095), 98 bis boulevard Arago, F-75014 Paris, France*
[2]*Institut de Physique Théorique, Orme des Merisiers, bâtiment 774, CEA/Saclay F-91191 Gif-sur-Yvette, France*
[3]*Department of Physics, University of Alberta, 11322-89 Avenue, Edmonton, Alberta, T6G 2G7, Canada*
[4]*Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan*

Accepted 2013 July 16. Received 2013 July 15; in original form 2013 May 31

ABSTRACT

In the context of upcoming large-scale structure surveys such as *Euclid*, it is of prime importance to quantify the effect of peculiar velocities on geometric probes. Hence, the formalism to compute *in redshift space* the geometrical and topological one-point statistics of mildly non-Gaussian 2D and 3D cosmic fields is developed. Leveraging the partial isotropy of the target statistics, the Gram–Charlier expansion of the joint probability distribution of the field and its derivatives is reformulated in terms of the corresponding anisotropic variables. In particular, the cosmic non-linear evolution of the Minkowski functionals, together with the statistics of extrema, is investigated in turn for 3D catalogues and 2D slabs. The amplitude of the non-Gaussian redshift distortion correction is estimated for these geometric probes. In 3D, gravitational perturbation theory is implemented in redshift space to predict the cosmic evolution of all relevant Gram–Charlier coefficients. Applications to the estimation of the cosmic parameters $\sigma(z)$ and $\beta = f/b_1$ from upcoming surveys are discussed. Such statistics are of interest for anisotropic fields beyond cosmology.

Key words: methods: analytical – galaxies: statistics – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

In modern cosmology, random fields (2D or 3D) are fundamental ingredients in the description of the large-scale matter density and the cosmic microwave background (CMB). The large-scale structure of the matter distribution in the Universe is believed to be the result of the gravitational growth of primordial nearly Gaussian small perturbations originating from quantum fluctuations. Deviations from Gaussianity inevitably arise due to the non-linear dynamics of the growing structures, but may also be present at small, but potentially detectable levels in the initial seed inhomogeneities. Thus, studying non-Gaussian (NG) signatures in the random fields of cosmological data provides methods to learn both the details of early Universe's physics and mechanisms for structure's growth, addressing issues such as the matter content of the Universe (Zunckel, Gott & Lunnan 2011), the role of bias between galaxies and dark matter (DM) distributions (Desjacques & Sheth 2010), and whether it is dark energy or a modification to Einstein's gravity (Wang, Chen & Park 2012) that is responsible for the acceleration of the Universe's expansion.

With upcoming high-precision surveys, it has become necessary to revisit alternative tools to investigate the statistics of random cosmological fields so as to handle observables with different sensitivity. Minkowski functionals (Mecke, Buchert & Wagner 1994) have been actively used (Gott, Weinberg & Melott 1987; Weinberg, Gott & Melott 1987; Melott, Weinberg & Gott 1988; Gott et al. 1989, 2007; Hikage et al. 2002, 2003; Park et al. 2005; Planck Collaboration 2013) as an alternative to the usual direct measurements of higher order moments and N-point correlation functions (Scoccimarro et al. 1998; Percival et al. 2007; Gaztañaga et al. 2009; Nishimichi et al. 2010, amongst many other studies) as they will present different biases and might be more robust, e.g. with regard to rare events. These functionals are topological and geometrical estimators involving the critical sets of the field. As such they are invariant under a monotonic transformation of the field (and in particular bias independent if the bias is monotonic!). If their expression for Gaussian isotropic fields has long been known (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Hamilton, Gott & Weinberg 1986; Bond & Efstathiou 1987; Ryden 1988; Ryden et al.

⋆E-mail: codis@iap.fr

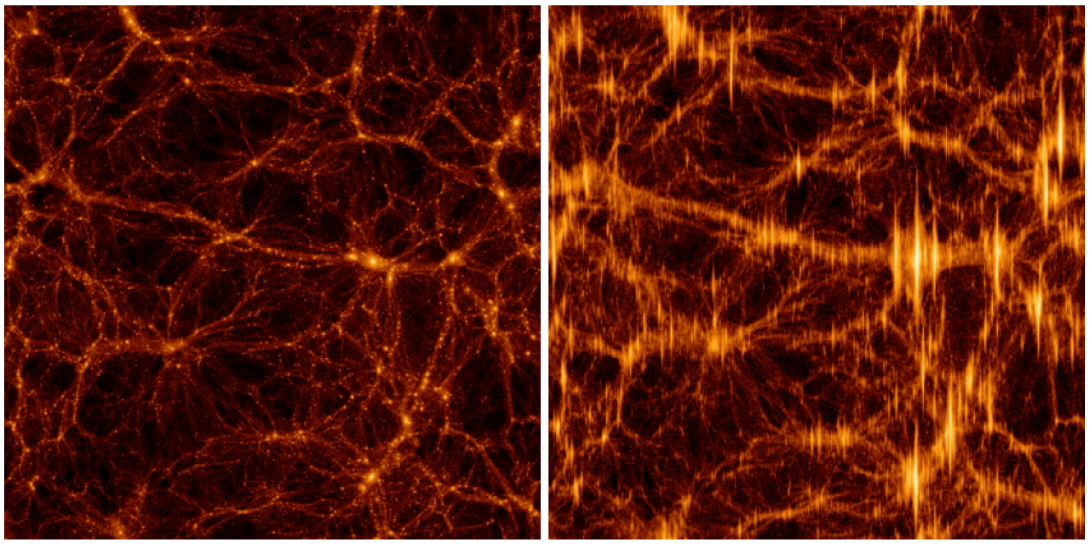


Figure 1. Left-hand panel: an example of a slice through 512^3 DM particles in ΛCDM simulations at redshift 0 ($\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.92$) in real space. The box size is 100 Mpc h^{-1} and the slice thickness is 10 Mpc h^{-1}. Right-hand panel: the same field when redshift distortion has been applied along the ordinate. Fingers of God are quite visible on that slice (see also Fig. 2).

1989), their theoretical prediction has also been computed more recently for mildly NG fields [Matsubara 1994; Pogosyan, Gay & Pichon 2009b; Matsubara 2010, for the first and second NG corrections using a multivariate Edgeworth expansion, and more recently by Gay, Pichon & Pogosyan (2012) to all orders in non-Gaussianity]. However, one major assumption in these results has been the isotropy of the underlying field.

Indeed, the assumptions of homogeneity and isotropy of our observable Universe are central to our understanding of the universe. This focuses our primary attention on statistically homogeneous and isotropic random fields as the description of cosmological 2D and 3D data. These statistical symmetries provide essential guidance for the theoretical description of the geometry of the cosmological random fields. Recent papers (Pogosyan et al. 2009b; Pogosyan, Pichon & Gay 2011; Gay et al. 2012) leverage the isotropy of the target statistics to develop NG moment expansion to all orders for popular and novel statistics such as the above-mentioned Minkowski functionals, but also extrema counts and the skeleton of the filamentary structures. This formalism, generalizing the earlier works (Bardeen et al. 1986; Matsubara 1994, 2003; Pogosyan et al. 2009a), allows for controlled comparison of the NG deviations of geometrical measures in theoretical models, simulations and the observations, assuming homogeneity and isotropy. Nevertheless, 3D surveys are conveyed in redshift space where the hypothesis of isotropy breaks down. Indeed in astrophysical observations, the three-dimensional positions of structures are frequently not accessible directly. While angular positions on the sky can be obtained precisely, the radial line-of-sight (hereafter LOS) position of objects is determined by proxy, e.g., via measurements of the LOS velocity component. This implies that galaxy distribution data are presented in redshift space in cosmological studies. Figs 1 and 2 illustrate the amplitude of such redshift distortion as a function of scale. On small, non-linear scales, the well-known finger-of-God effect (Jackson 1972; Peebles 1980) stretches the collapsing clusters along the LOS whereas on larger scales (see Fig. 2) the redshift-space distortion flattens the voids along the LOS (Sargent & Turner 1977), in accordance with the linear result of Kaiser (1987). But anisotropy in data does not affect only cosmological surveys: this issue is ubiquitous in physics: e.g. the neutral hydrogen distribution is mapped in position–position–velocity cubes in studies of the interstellar or intergalactic medium, and the turbulent distribution of magnetic field is mapped via Faraday rotation measure of the synchrotron emission (Heyvaerts et al. 2013). Thus, as a rule, the underlying isotropy of structures in real space is broken in the space where data are available. Hence, developing techniques to recover the properties of underlying fields from distorted data sets, anisotropic in the LOS direction, is a fundamental problem in astrophysics.

In this paper, we thus present a theory of the NG Minkowski functionals and extrema counts for z-anisotropic cosmological fields, targeted for application to data sets in redshift space in the so-called plane-parallel approximation. Hence, this theory is a generalization of the formalism for mildly NG but isotropic fields of Pogosyan et al. (2009b) and Gay et al. (2012) on one hand, and the theory of anisotropic redshift-space effects on statistics in the Gaussian limit developed by Matsubara (1996) on the other hand. The effects of anisotropy and non-Gaussianity are simultaneously important for a precise description of the large-scale structures (LSS hereafter) of the Universe as mentioned in Matsubara & Suto (1995), where N-body simulations suggest that redshift-space distortion has noticeable impact on the shape of the genus curve in the weakly non-linear regime. At the same time, the theory presented in this paper is general and applicable to mildly NG homogeneous and statistically axisymmetric random fields of any origin, for example for extending velocity channel analysis of H I maps (Lazarian & Pogosyan 2000) to account for NG density compressibility, or describing cosmological perturbations in anisotropic Bianchi models of the Universe.

As an anticipation, Fig. 3 illustrates the importance of modelling appropriately the anisotropy on the particular example of the 3D Euler characteristics of a mildly NG scale-invariant cosmological field. Indeed, the theoretical prediction assuming isotropy given in equation (38)

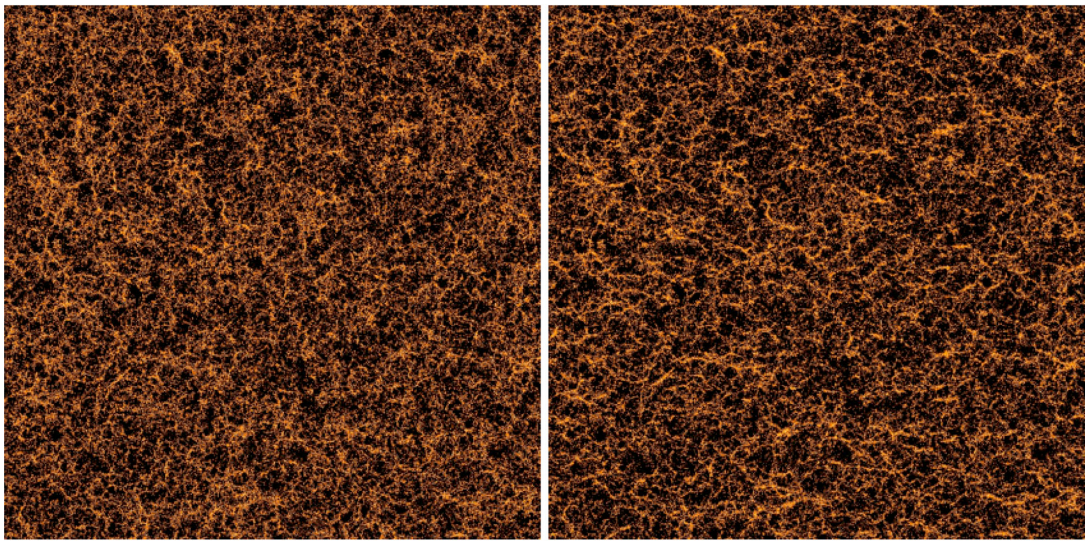


Figure 2. A slice through the HORIZON 4π halo catalogue at redshift zero without (left-hand panel) and with (right-hand panel) redshift distortion (along the ordinate). The box size is 2 Gpc h^{-1} and the slice thickness is 40 Mpc h^{-1}. The area is 20×20 larger than the slice presented in Fig. 1. Each dot represents a halo colour coded by its mass. Note the clear preferred horizontal elongation of structures in redshift space.

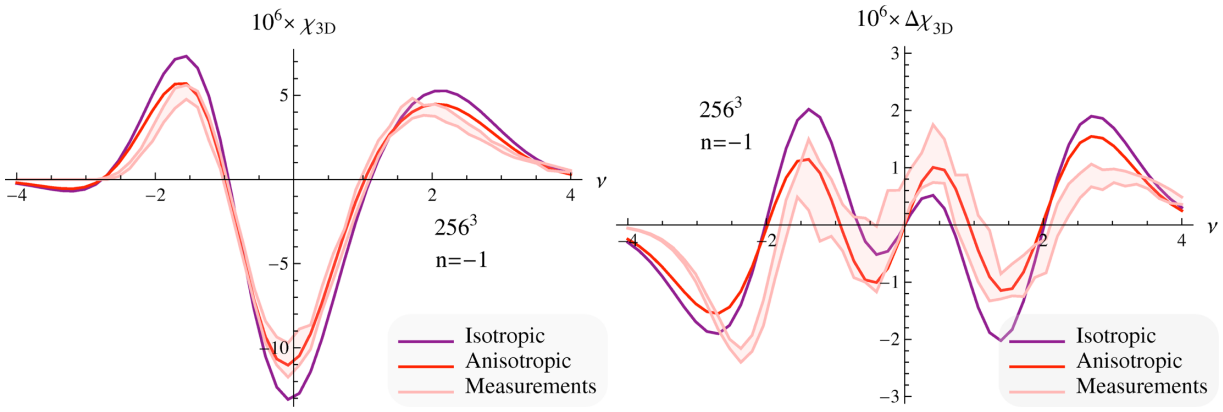


Figure 3. Left-hand panel: 3D Euler characteristic of a mildly NG scale-invariant cosmological field ($n = -1, f = 1, \sigma = 0.18$). The prediction to first order in non-Gaussianity given by equation (35) corresponds to the red solid line and the prediction assuming isotropy (see equation 38) to the purple line. The one-sigma shaded area is the measurement. Right-hand panel: same as the left-hand panel for the first-order correction to the Gaussian distribution only.

below (in purple) significantly fails to describe the measurement unlike the prediction proposed in equation (35) (in red). This suggests that in redshift space it is of great importance to properly account for anisotropy. This will be the topic of this paper. Section 2 defines the statistics used in this paper, namely Minkowski functionals and critical point counts. Section 3 introduces the formalism to deal with the joint probability density function (JPDF) of the density field and its derivatives up to second order in redshift space. In particular, it shows how rotational invariance is used to introduce a relevant set of variables which diagonalizes the Gaussian JPDF; a Gram–Charlier (GC) expansion of the NG JPDF is written there. Section 4 presents Minkowski functionals and extrema counts in 2D and 3D. Section 4.2 presents the full NG expression for 2D and 3D Euler characteristic in redshift space, while Section 4.3 is devoted to the last three Minkowski functionals (area of isosurfaces, length of isocontours and contour crossing) for which we present expressions in redshift space up to first order in non-Gaussianity. Section 4.4 sketches the derivation for extrema counts in two and three dimensions. Section 4.5.1 re-expresses these functionals as a function of the filling factor threshold, while Section 4.5.2 investigates the implications of the topological invariance for the corresponding set of cumulants of the field. Section 5 analyses implication for the estimation of cosmological parameters. It shows how to compute the relevant three-point cumulants in redshift space (e.g. skewness and its generalization to the derivatives of the density field) at tree order within perturbation theory (PT). In particular, it provides a way to compute analytically the angular part of the integrals under consideration. It then discusses which features of the bispectrum are robustly measured using these NG critical sets and sketches two main applications: measuring σ (hence dark energy) of the possibly masked underlying field and measuring $\beta \equiv f/b_1 \simeq \Omega_{\rm m}^{0.55}/b_1$. Appendix A lists the properties of the relevant cumulants. Appendix B presents briefly a set of '$f_{\rm nl}$' anisotropic NG field toy models which we use to validate our theory. Finally Appendix C derives the 3D Euler characteristic at all orders in non-Gaussianity.

2 GEOMETRICAL STATISTICS IN 2D AND 3D

Our prime focus is the geometrical and topological properties of cosmological fields. One way to probe these is to look at isocontours of the field at different thresholds and use Minkowski functionals to describe them (Mecke et al. 1994). These Minkowski functionals are known to be the only morphological descriptors in integral geometry that respect motion invariance, conditional continuity and additivity (Hadwiger 1957). As a result, they form a robust and meaningful set of observables. In d dimensions, there are $d + 1$ such functionals (four in 3D and three in 2D), namely in 3D: the encompassed volume, f_V, the surface area, $\mathcal{N}_3$, the integral mean curvature and the integral Gaussian curvature (closely related to the Euler–Poincaré characteristic, χ). For random fields these functionals are understood as densities, i.e. quantities per unit volume of space.

Especially when studying the anisotropic field, complimentary information can be obtained by using the geometrical statistics and Minkowski functionals for the field obtained on lower dimensional sections of the 3D field. For example, in addition to 3D isocontour area statistics, one can introduce the length of 2D isocontours on a planar section $\mathcal{N}_2$, and contour crossings by a line through 3D space, $\mathcal{N}_1$. These statistics for cosmology were first introduced by Ryden (1988) and Ryden et al. (1989). In the isotropic limit, they are trivially related: $2\mathcal{N}_1 = 4\mathcal{N}_2/\pi = \mathcal{N}_3$; this relation does not hold anymore for an anisotropic field as it will be shown in Sections 4.3.1, 4.3.2 and 4.3.3. Similarly, in addition to the full Euler characteristic $\chi_{3\mathrm{D}}$ of 3D excursion sets, we shall consider the 2D Euler characteristic, $\chi_{2\mathrm{D}}$, on planar sections through the field.

All geometrical measures can be expressed as averages over the JPDF of the field and its derivatives. In the following, let us call x the field under consideration and, without loss of generality, assume that it has zero mean. In cosmological applications, this field, for instance, can be the 3D density contrast. Collecting the well-known results from an extensive literature (e.g. Rice 1944, 1945; Ryden 1988; Matsubara 1996) in a compact form, we have for the first two Minkowski functionals

$$f_V(\nu) = \langle \Theta(x-\nu) \rangle \,, \tag{1}$$

$$\mathcal{N}_3(\nu) = \langle |\nabla x| \delta_{\mathrm{D}}(x-\nu) \rangle \,, \quad \mathcal{N}_2(\nu) = \langle |\nabla_{\mathcal{S}} x| \delta_{\mathrm{D}}(x-\nu) \rangle \,, \quad \mathcal{N}_1(\nu) = \langle |\nabla_{\mathcal{L}} x| \delta_{\mathrm{D}}(x-\nu) \rangle \,, \tag{2}$$

where the δ_{D}-function in the statistical averaging signifies evaluation at the given threshold $x = \nu$, while the step function reflects the cumulative averaging over the values above the threshold $x \geq \nu$. We see that the family of threshold-crossing statistics is given by the average gradient of the field for $\mathcal{N}_3$ or its restriction to the plane $\mathcal{S}$ or line $\mathcal{L}$ for $\mathcal{N}_2$ and $\mathcal{N}_1$, respectively.

The average of the Gaussian curvature on the isosurface is, via the Gauss–Bonnet theorem, its topological Euler characteristic χ, which thus can be expressed directly as $\chi(\nu) = \langle \delta_{\mathrm{D}}(x-\nu)\delta_{\mathrm{D}}(\nabla_1 x)\delta_{\mathrm{D}}(\nabla_2 x)|\nabla_3 x|(\nabla_1\nabla_1 x\nabla_2\nabla_2 x - (\nabla_1\nabla_2 x)^2)\rangle$ (Hamilton et al. 1986; Matsubara 1996). In this paper, we use the Euler characteristic, $\chi_{3\mathrm{D}}$, of the excursion set encompassed by the isosurface (which in 3D is just one half of the Euler characteristic of the isosurface itself, and is equal to minus the genus for the definitions used in cosmology; see detailed discussions for such conventions in Gay et al. 2012). Being the alternating sum of Betti numbers, $\chi_{3\mathrm{D}}$ is related via Morse theory (e.g. Jost 2008) to the alternating sum of the number of critical points in the excursion volume. For a random field (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986)[1]

$$\chi_{3\mathrm{D}}(\nu) = -\left\langle \det\left(\nabla_i \nabla_j x\right) \delta_{\mathrm{D}}(\nabla x)\Theta(x-\nu)\right\rangle, \quad i,j \in \{1,2,3\}. \tag{3}$$

The Euler characteristic of the excursion sets of a 2D field (in particular, 2D slices of a 3D random field) is given by a similar expression (Adler 1981; Bond & Efstathiou 1987; Coles 1988; Melott et al. 1989; Gott et al. 1990)

$$\chi_{2\mathrm{D}}(\nu) = \left\langle \det\left(\nabla_i \nabla_j x\right) \delta_{\mathrm{D}}(\nabla x)\Theta(x-\nu)\right\rangle \,, \quad i,j \in \{1,2\} \,. \tag{4}$$

For an anisotropic 3D field along the third direction, $\chi_{2\mathrm{D}}$ depends on the angle $\theta_{\mathcal{S}}$ between this direction and the plane ($\mathcal{S}$) under consideration (Matsubara 1996). Equation (4) is then understood as

$$\chi_{2\mathrm{D}}(\nu,\theta_{\mathcal{S}}) = \left\langle \det\left(\tilde{\nabla}_i \tilde{\nabla}_j x\right) \delta_{\mathrm{D}}(\tilde{\nabla} x)\Theta(x-\nu)\right\rangle \,, \quad i,j \in \{1,2\}, \tag{5}$$

where $\tilde{\nabla}$ is the gradient on the plane, i.e. $\tilde{\nabla}_1 = \nabla_1$ and $\tilde{\nabla}_2 = \sin\theta_{\mathcal{S}}\nabla_2 + \cos\theta_{\mathcal{S}}\nabla_3$. The freedom associated with the choice of plane will be further discussed in Section 5.

With the same formalism, it is easy to compute the critical point counts (Adler 1981; Bardeen et al. 1986). Equation (3) leads to a cumulative counting above a given threshold for maxima, two types (filamentary and wall-like) of saddle points, and minima

$$n_{\mathrm{max,3D}}(\nu) = -\left\langle \det\left(\nabla_i \nabla_j x\right) \delta_{\mathrm{D}}(\nabla x)\Theta(-\lambda_1)\Theta(x-\nu)\right\rangle, \tag{6}$$

$$n_{\mathrm{sadf,3D}}(\nu) = +\left\langle \det\left(\nabla_i \nabla_j x\right) \delta_{\mathrm{D}}(\nabla x)\Theta(\lambda_1)\Theta(-\lambda_2)\Theta(x-\nu)\right\rangle, \tag{7}$$

$$n_{\mathrm{sadw,3D}}(\nu) = -\left\langle \det\left(\nabla_i \nabla_j x\right) \delta_{\mathrm{D}}(\nabla x)\Theta(\lambda_2)\Theta(-\lambda_3)\Theta(x-\nu)\right\rangle, \tag{8}$$

[1] Note that this expression comes from $\delta_{\mathrm{D}}(\nabla x) = \sum_{x_0|\nabla x_0=0} \delta_{\mathrm{D}}(x-x_0)/|\det\left(\nabla_i\nabla_j x\right)|$ and the absolute value of the Hessian can be dropped because we are interested in the *alternating* sum of critical points.

$$n_{\min,3D}(\nu) = + \left\langle \det\left(\nabla_i \nabla_j x\right) \delta_D(\nabla x)\Theta(\lambda_3)\Theta(x-\nu)\right\rangle \ , \tag{9}$$

where averaging conditions are set by the signs of sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of the Hessian matrix of the field. Taking alternating sum eliminates the constraints on signs of eigenvalue, leading to the χ_{3D} statistics. Similar expressions as equations (6)–(9) apply for 2D extrema.

Extrema counts provide us with information on peaks (dense regions), minima (underdense regions) and saddle points. In some applications, there is symmetry between extrema (e.g. in CMB studies, minima and maxima of the temperature field are equivalent); in others, they describe very different structures, e.g. in LSS dense peaks correspond to gravitationally collapsing objects such as galactic or cluster haloes while minima seed the regions devoid of structures. Saddle-type extrema are also interesting in their own right, being related to the underlying filamentary structures (bridges connecting peaks through saddles), which in turn can also be characterized by the skeleton (Novikov, Colombi & Doré 2003; Gay et al. 2012) of the cosmic web. A particular advantage of the described geometrical statistical estimates is that they are invariant under the monotonic transformation of the underlying field $x \to f(x)$, provided one maps the threshold correspondingly $\nu \to f(\nu)$. For cosmological data, this means that these statistics are formally invariant with respect to any monotonic local bias between the galaxy and matter distributions. We demonstrate this formally, order by order, in Section 4.5.2.

3 THE JPDF: ROTATIONAL INVARIANCE AND GRAM–CHARLIER EXPANSION

Evaluating the expectations in equations (1)–(9) requires a model for the JPDF of the field and its derivative up to second order, $P(x, \nabla_i x, \nabla_i \nabla_j x)$. Let us now proceed to developing this JPDF for a mildly NG and anisotropic field such as the cosmological density field in redshift space, starting with a formal definition of redshift space.

3.1 Statistically anisotropic density field in redshift space

In an astrophysical context, we focus on the statistics of isodensity contours of matter in the redshifted Universe. The estimation of position via redshift assigns to a given object the 'redshift' coordinate, $\boldsymbol{s}$,

$$\boldsymbol{s} = \boldsymbol{r} + H^{-1}\boldsymbol{v}\cdot\hat{\boldsymbol{r}} \tag{10}$$

shifted from the true position $\boldsymbol{r}$ by the projection of the peculiar velocity $\boldsymbol{v}$ along the LOS direction $\hat{\boldsymbol{r}}$.

On large scales, in the linear regime of density evolution, the mapping to redshift coordinates induces an anisotropic change in mass density contrast (Kaiser 1987), best given in Fourier space

$$\hat{\delta}^{(s)}(\boldsymbol{k}) = (1 + f\mu^2)\hat{\delta}^{(r)}(\boldsymbol{k}), \tag{11}$$

that has dependence on the angle $\mu = \boldsymbol{k}\cdot\hat{\boldsymbol{r}}/k$ between the direction of the wave $\boldsymbol{k}$ and the LOS, and the amplitude, f, tracing the growth history of linear inhomogeneities $D(a)$, $f = \mathrm{d}\log D/\mathrm{d}\log a \approx \Omega_{\rm m}^{0.55}$ (Peebles 1980). The main qualitative effect of this distortion is the enhancement of clustering via the squeezing overdense regions and the stretching underdense voids along the LOS. If matter is traced by biased haloes (e.g. galaxies), the redshift-space distortion δ_g can be modelled by a linear bias b_1 factor (Kaiser 1984),

$$\hat{\delta}_g^{(s)}(\boldsymbol{k}) = (1 + \beta\mu^2)\hat{\delta}_g^{(r)}(\boldsymbol{k}), \quad \beta \equiv f/b_1. \tag{12}$$

Depending on context, we shall use either f or β to parametrize the linearized redshift distortions, and either equation (11) or (12) to generate maps.

In the mildly non-linear regime, the focus of this investigation, redshift distortions interplay with NG corrections that develop with the growth of non-linearities. Scoccimarro, Couchman & Frieman (1999) and Bernardeau et al. (2002) established the framework for a perturbative approach to this regime, which we built upon in Section 5.3. In redshift space, the density field is statistically anisotropic, with LOS expectations differing from expectations in the perpendicular directions (in the plane of the sky). We shall consider the Minkowski functionals and extrema statistics in the plane-parallel approximation, where the LOS direction is identified with the Cartesian third, 'z', coordinate. As the first step, we extend the NG formalism introduced in Pogosyan et al. (2009b) and Gay et al. (2012) to partly anisotropic, axisymmetric fields, establishing several formal results for all orders in non-Gaussianity.

3.2 3D formalism

Following Pogosyan et al. (2009b), we choose an NG GC expansion of the JPDF (Cramér 1946; Kendall & Stuart 1958; Chambers 1967; Juszkiewicz et al. 1995; Amendola 1996; Blinnikov & Moessner 1998) (for a first application to CMB, see, e.g., Scaramella & Vittorio 1991) using polynomial variables that are invariant with respect to the statistical symmetries of the field. In the presence of anisotropy in the direction along the LOS, all statistical measures should be independent with respect to sky rotations in the plane perpendicular to the LOS.

Let us denote the field variable as x for the density contrast and x_i, x_{ij} for its first and second derivatives. These field variables are separated into the following groups based on their behaviour under sky rotation: x and x_{33} are scalars, x_3 is a pseudo-scalar, (x_1, x_2) and (x_{13}, x_{23}) are two vectors and (x_{11}, x_{22}, x_{12}) is a symmetric 2×2 tensor. We can construct eight 2D sky rotation invariant polynomial quantities: four linear ones x, x_{33}, x_3 and $J_{1\perp} \equiv x_{11} + x_{22}$; three quadratic: $q_\perp^2 \equiv x_1^2 + x_2^2$, $Q^2 \equiv x_{13}^2 + x_{23}^2$ and $J_{2\perp} \equiv (x_{11} - x_{22})^2 + 4x_{12}^2$; and one cubic,

$\Upsilon \equiv (x_{13}^2 - x_{23}^2)(x_{11} - x_{22}) + 4x_{12}x_{13}x_{23}$, which is directly related to the 2D polar angle ψ between the (x_{13}, x_{23}) vector and the eigendirection of the (x_{11}, x_{22}, x_{12}) matrix,[2] $\Upsilon = Q^2\sqrt{J_{2\perp}}\cos 2\psi$.

To build the GC expansions, we start with the Gaussian limit to the one-point JPDF of all the invariant variables, which then serves as a kernel for defining orthogonal polynomials into which the deviations from Gaussianity are expanded (see Gay et al. 2012 for further details). The Gaussian limit is determined as the limit in which the distribution of field variables x, x_i, x_{ij} is approximated by the Gaussian JPDF. The field variables are defined to have zero mean. Their covariance matrix in the anisotropic case contains the variances, defined as

$$\left\langle x^2\right\rangle \equiv \sigma^2, \quad \left\langle x_1^2\right\rangle = \left\langle x_2^2\right\rangle \equiv \frac{1}{2}\sigma_{1\perp}^2, \quad \left\langle x_3^2\right\rangle \equiv \sigma_{1\parallel}^2, \quad \left\langle x_{11}^2\right\rangle = \left\langle x_{22}^2\right\rangle \equiv \frac{3}{8}\sigma_{2\perp}^2,$$

$$\left\langle x_{12}^2\right\rangle \equiv \frac{1}{8}\sigma_{2\perp}^2, \quad \left\langle x_{33}^2\right\rangle \equiv \sigma_{2\parallel}^2, \quad \left\langle x_{13}^2\right\rangle = \left\langle x_{23}^2\right\rangle \equiv \frac{1}{2}\sigma_Q^2, \tag{13}$$

and the cross-correlations, amongst which the non-zero are

$$\langle x x_{11}\rangle = \langle x x_{22}\rangle \equiv -\frac{1}{2}\sigma_{1\perp}^2, \quad \langle x x_{33}\rangle \equiv -\sigma_{1\parallel}^2, \quad \langle x_{11}x_{22}\rangle \equiv \frac{1}{8}\sigma_{2\perp}^2, \quad \langle x_{11}x_{33}\rangle = \langle x_{22}x_{33}\rangle = \frac{1}{2}\sigma_Q^2. \tag{14}$$

These properties translate into the following lowest moments for 2D rotation invariant quantities

$$\langle x\rangle = \langle J_{1\perp}\rangle = \langle x_{33}\rangle = \langle\Upsilon\rangle = 0, \quad \left\langle x^2\right\rangle = \sigma^2, \quad \langle J_{1\perp}^2\rangle = \sigma_{2\perp}^2, \quad \left\langle x_{33}^2\right\rangle = \sigma_{2\parallel}^2, \quad \langle q_\perp^2\rangle = \sigma_{1\perp}^2, \quad \left\langle Q^2\right\rangle = \sigma_Q^2, \quad \langle J_{2\perp}\rangle = \sigma_{2\perp}^2. \tag{15}$$

The cross-correlations of invariants are limited to the x, $J_{1\perp}$, x_{33} subset, $\langle x J_{1\perp}\rangle = -\sigma_{1\perp}^2$, $\langle x x_{33}\rangle = -\sigma_{1\parallel}^2$, $\langle J_{1\perp}x_{33}\rangle = \sigma_Q^2$ and the coupling of Υ with Q^2 and $J_{2\perp}$. The scalar part of the gradient x_3 remains uncorrelated with the rest of the variables due to its pseudo-scalar nature (changing sign when z flips). From now on, we shall consider field variables to be normalized by the corresponding σ, $\sigma_{1\perp}$, $\sigma_{1\parallel}$, $\sigma_{2\perp}$, $\sigma_{2\parallel}$ and σ_Q. The linear invariant combinations are normalized by their standard deviations and the quadratic ones by their mean values. Correlations are then described by the dimensionless coefficients $\gamma_\perp \equiv \sigma_{1\perp}^2/(\sigma\sigma_{2\perp})$, $\gamma_\parallel \equiv \sigma_{1\parallel}^2/(\sigma\sigma_{2\parallel})$ and $\gamma_2 \equiv \sigma_Q^2/(\sigma_{2\perp}\sigma_{2\parallel})$, corresponding to generalized shape parameters (Bernardeau et al. 2002). Note that in the isotropic limit,

$$\sigma_{1\perp}^2 = 2/3\sigma_1^2, \quad \sigma_{1\parallel}^2 = 1/3\sigma_1^2, \quad \sigma_{2\perp}^2 = 8/15\sigma_2^2, \quad \sigma_{2\parallel}^2 = 1/5\sigma_2^2, \quad \sigma_Q^2 = 2/15\sigma_2^2, \quad \gamma_\perp = \sqrt{5/6}\gamma, \quad \gamma_\parallel = \sqrt{5/9}\gamma, \quad \gamma_2 = 1/\sqrt{6}, \tag{16}$$

where $\sigma_1^2 \equiv \langle x_1^2 + x_2^2 + x_3^2\rangle = \langle\nabla x\cdot\nabla x\rangle$, $\sigma_2^2 = \langle(x_{11} + x_{22} + x_{33})^2\rangle = \langle(\Delta x)^2\rangle$ and $\gamma \equiv \sigma_1^2/(\sigma\sigma_2)$.

Let us now also introduce a decorrelated set of invariant variables (x, ξ, ζ) with the following combinations of x, $J_{1\perp}$, x_{33}:

$$x_{33} = \xi\sqrt{\frac{1-\gamma_\perp^2-\gamma_\parallel^2-\gamma_2^2+2\gamma_\perp\gamma_\parallel\gamma_2}{1-\gamma_\perp^2}} + \frac{\gamma_2-\gamma_\parallel\gamma_\perp}{\sqrt{1-\gamma_\perp^2}}\zeta - \gamma_\parallel x\ , \quad J_{1\perp} = \zeta\sqrt{1-\gamma_\perp^2} - \gamma_\perp x. \tag{17}$$

The resulting Gaussian distribution $G = G(x, q_\perp^2, x_3, \zeta, J_{2\perp}, \xi, Q^2, \Upsilon)$ then simply reads in terms of these variables

$$G\,\mathrm{d}x\,\mathrm{d}q_\perp^2\mathrm{d}x_3\,\mathrm{d}\zeta\,\mathrm{d}J_{2\perp}\mathrm{d}\xi\,\mathrm{d}Q^2\mathrm{d}\Upsilon = \frac{1}{4\pi^3}\mathrm{e}^{-\frac{1}{2}x^2-q_\perp^2-\frac{1}{2}x_3^2-\frac{1}{2}\zeta^2-J_{2\perp}-\frac{1}{2}\xi^2-Q^2}\mathrm{d}x\,\mathrm{d}q_\perp^2\mathrm{d}x_3\,\mathrm{d}\zeta\,\mathrm{d}J_{2\perp}\mathrm{d}\xi\,\mathrm{d}Q^2\frac{d\Upsilon}{\sqrt{Q^4J_{2\perp}-\Upsilon^2}}, \tag{18}$$

where x, x_3, ξ, ζ vary in the range $]-\infty,\infty[$, $q_\perp^2$, Q^2, $J_{2\perp}$ span positive values $[0,\infty[$ and Υ is limited to $[-Q^2\sqrt{J_{2\perp}}, Q^2\sqrt{J_{2\perp}}]$.

The GC polynomial expansion for an NG JPDF is then obtained by using polynomials that are orthogonal with respect to the kernel provided by equation (18). The remaining coupling between Υ and the $(Q^2, J_{2\perp})$ variables in equation (18) introduces a technical complexity in building an explicit set of such polynomials. Fortunately, for most of the geometrical statistics considered in this paper, the Υ dependence is trivial and we can limit ourselves to probability density functions (PDFs) marginalized over Υ. After Υ marginalization, all the remaining variables are uncorrelated in the Gaussian limit, and the NG JPDF, $P(x, q_\perp^2, x_3, \zeta, J_{2\perp}, \xi, Q^2)$, can be expanded in a series of direct products of the familiar Hermite (for which we use the 'probabilists' convention) and Laguerre polynomials

$$P = G\left[1 + \sum_{n=3}^{\infty}\sum_{\sigma_n}\frac{(-1)^{j+l+r}}{i!\,j!\,k!\,l!\,m!\,r!\,p!}\left\langle x^i q_\perp^{2j}\zeta^k J_{2\perp}^l x_3^m Q^{2r}\xi^p\right\rangle_{\mathrm{GC}} H_i(x)L_j\left(q_\perp^2\right)H_k(\zeta)L_l(J_{2\perp})H_m(x_3)L_r\left(Q^2\right)H_p(\xi)\right], \tag{19}$$

where $\sum_{\sigma_n}$ is the sum over all combinations of indices $\sigma_n = \{(i, j, k, l, m, p, r) \in \mathbb{N}^7 | i + 2j + k + 2l + m + 2r + p = n\}$ such that powers of the field add up to n, and G is given by equation (18) after integration over Υ. The terms within the expansion (19) are sorted in the order of the power in the field variable n. The GC coefficients are defined by

$$\left\langle x^i q_\perp^{2j}\zeta^k J_{2\perp}^l x_3^m Q^{2r}\xi^p\right\rangle_{\mathrm{GC}} \equiv (-1)^{j+l+r}\,j!\,l!\,r!\left\langle H_i(x)L_j(q_\perp^2)H_k(\zeta)L_l(J_{2\perp})H_m(x_3)L_r(Q^2)H_p(\xi)\right\rangle, \tag{20}$$

normalized so that $\langle x^i q_\perp^{2j}\zeta^k J_{2\perp}^l x_3^m Q^{2r}\xi^p\rangle_{\mathrm{GC}} = \langle x^i q_\perp^{2j}\zeta^k J_{2\perp}^l x_3^m Q^{2r}\xi^p\rangle$ + products of lower order moments. The advantage of using strictly polynomial variables is that all the moments that appear in the GC coefficients can be readily related to the moments of the underlying field, and can be obtained if the theory of the latter (for example, PT of gravitational instability) is known. As shown in Gay et al. (2012), at the

[2] There is no quadratic combination that would represent this angle via a scalar product of two vectors. The reason is that any 'vector' built linearly from x_{ij}, such as the 'Q,U' one $(x_{11} - x_{22}, 2x_{12})$, rotates with twice the rotation angle when the real vectors, e.g. (x_{13}, x_{23}), rotate normally. The combination Υ can be seen to be a scalar product of 'vectors' , e.g. $(x_{11} - x_{22}, 2x_{12})$ and $(x_{13}^2 - x_{23}^2, 2x_{13}x_{23})$.

lowest NG ($n = 3$) order, the GC coefficients are just equal to the moments of the corresponding variables, while in the next two orders they coincide with their cumulants, defined as 'field cumulants'[3] if the variable is non-linear (for details, see Gay et al. 2012).

Expression (19) is somewhat simplified under the condition of zero gradient, arising, e.g., when investigating the Euler characteristic and extrema densities,

$$P_{\rm ext}(x,\zeta,J_{2\perp},\xi,Q^2)=G_{\rm ext}\left[1+\sum_{n=3}^{\infty}\sum_{\sigma_n}\frac{(-1)^{j+l+r+m}2^{-m}}{i!\,j!\,k!\,l!\,m!\,r!\,p!}\left\langle x^i q_\perp^{2j}\zeta^k J_{2\perp}^l x_3^{2m}Q^{2r}\xi^p\right\rangle_{\rm GC}H_i(x)H_k(\zeta)L_l(J_{2\perp})L_r\left(Q^2\right)H_p(\xi)\right], \tag{21}$$

since $L_j(0)=1$, $H_{2m}(0)=(-1)^m(2m)!/(2^m m!)$ and $H_{2m+1}(0)=0$; $G_{\rm ext}=\frac{1}{\pi}G(x,0,0,\zeta,J_{2\perp},\xi,Q^2)$, where the $1/\pi$ comes from the use of polar ($q_\perp^2$) versus Cartesian coordinates (x_1,x_2). Note that m is then replaced by $2m$ in the power count in σ_n.

3.3 Theory on 2D planes

One of our purposes is to study Minkowski functionals on 2D planar sections of 3D fields. Let us therefore introduce the anisotropic 2D (on the plane) formalism. In a 2D planar slice through anisotropic space no residual symmetries are left, so we may use the field variables directly. The emphasis then is on relating the field properties on the plane to the three-dimensional ones. Denoting the basis vectors of 3D space as $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$, with $\boldsymbol{u}_3$ directed along the LOS, let us introduce the coordinate system on the 2D plane using the pair of basis vectors such that $\boldsymbol{s}_1=\boldsymbol{u}_1$ is perpendicular to the LOS and $\boldsymbol{s}_2=\cos\theta_{\mathcal{S}}\boldsymbol{u}_3+\sin\theta_{\mathcal{S}}\boldsymbol{u}_2$, where $\theta_{\mathcal{S}}$ is the angle between the LOS and that plane. We label the field variables on the plane with tilde, using the set $(\tilde{x},\tilde{x}_1,\tilde{x}_2,\tilde{x}_{11},\tilde{x}_{22},\tilde{x}_{12})$, where the relation with 3D spatial derivatives is established by $\partial_{s_1}=\partial_1$ and $\partial_{s_2}=\cos\theta_{\mathcal{S}}\partial_3+\sin\theta_{\mathcal{S}}\partial_2$. Hereafter, the first direction corresponds to s_1 and the second direction to s_2. Note that $\tilde{x}$, $\tilde{x}_1$ and $\tilde{x}_{11}$ coincide with their 3D counterparts. We use variables rescaled by their respective variance denoted, using the same notation as in the 3D case, as $(\tilde{\sigma},\tilde{\sigma}_{1\perp},\tilde{\sigma}_{1\parallel},\tilde{\sigma}_{2\perp},\tilde{\sigma}_{2\parallel},\tilde{\sigma}_Q)$. These variances involve the plane orientation in the following way: $\tilde{\sigma}=\sigma$, $\tilde{\sigma}_{1\perp}^2=\langle x_1^2\rangle=\frac{1}{2}\sigma_{1\perp}^2$, $\tilde{\sigma}_{1\parallel}^2=\langle\tilde{x}_2^2\rangle=\cos^2\theta_{\mathcal{S}}\sigma_{1\parallel}^2+\frac{1}{2}\sin^2\theta_{\mathcal{S}}\sigma_{1\perp}^2$, $\tilde{\sigma}_{2\perp}^2=\langle x_{11}^2\rangle=\frac{3}{8}\sigma_{2\perp}^2$, $\tilde{\sigma}_{2\parallel}^2=\langle\tilde{x}_{22}^2\rangle=\cos^4\theta_{\mathcal{S}}\sigma_{2\parallel}^2+\frac{3}{8}\sin^4\theta_{\mathcal{S}}\sigma_{2\perp}^2+3\cos^2\theta_{\mathcal{S}}\sin^2\theta_{\mathcal{S}}\sigma_Q^2$ and $\tilde{\sigma}_Q^2=\langle\tilde{x}_{12}^2\rangle=\frac{1}{2}\cos^2\theta_{\mathcal{S}}\sigma_Q^2+\frac{1}{8}\sin^2\theta_{\mathcal{S}}\sigma_{2\perp}^2$. As previously, in order to diagonalize the JPDF, we introduce $\tilde{\xi}$ and $\tilde{\zeta}$ such that

$$\tilde{x}_{22}=\sqrt{\frac{1-\tilde{\gamma}_\perp^2-\tilde{\gamma}_\parallel^2-\tilde{\gamma}_2^2+2\tilde{\gamma}_\perp\tilde{\gamma}_\parallel\tilde{\gamma}_2}{1-\tilde{\gamma}_\perp^2}}\tilde{\xi}-\frac{\tilde{\gamma}_\parallel\tilde{\gamma}_\perp-\tilde{\gamma}_2}{\sqrt{1-\tilde{\gamma}_\perp^2}}\tilde{\zeta}-\tilde{\gamma}_\parallel x\,,\quad \tilde{x}_{11}=\sqrt{1-\tilde{\gamma}_\perp^2}\tilde{\zeta}-\tilde{\gamma}_\perp x\,, \tag{22}$$

where $\tilde{\gamma}_\perp=\tilde{\sigma}_{1\perp}^2/(\tilde{\sigma}\tilde{\sigma}_{2\perp})$, $\tilde{\gamma}_\parallel=\tilde{\sigma}_{1\parallel}^2/(\tilde{\sigma}\tilde{\sigma}_{2\parallel})$ and $\tilde{\gamma}_2=\tilde{\sigma}_Q^2/(\tilde{\sigma}_{2\parallel}\tilde{\sigma}_{2\perp})$. In terms of these variables, the resulting Gaussian distribution is simply

$$G_{\rm 2D}(\tilde{x},\tilde{x}_1,\tilde{x}_2,\tilde{\zeta},\tilde{\xi},\tilde{x}_{12})=\frac{1}{8\pi^3}\exp\left[-\frac{1}{2}\tilde{x}^2-\frac{1}{2}\tilde{x}_1^2-\frac{1}{2}\tilde{x}_2^2-\frac{1}{2}\tilde{\zeta}^2-\frac{1}{2}\tilde{\xi}^2-\frac{1}{2}\tilde{x}_{12}^2\right], \tag{23}$$

and the fully NG JPDF can be written using a GC expansion in Hermite polynomials only,

$$P_{\rm 2D}(\tilde{x},\tilde{x}_1,\tilde{x}_2,\tilde{\zeta},\tilde{\xi},\tilde{x}_{12})=G_{\rm 2D}\left[1+\sum_{n=3}^{\infty}\sum_{\sigma_n}\frac{1}{i!\,j!\,k!\,l!\,m!\,p!}\left\langle\tilde{x}^i\tilde{x}_1^j\tilde{x}_2^k\tilde{\zeta}^l\tilde{\xi}^m\tilde{x}_{12}^p\right\rangle_{\rm GC}H_i(\tilde{x})\,H_j(\tilde{x}_1)H_k(\tilde{x}_2)H_l(\tilde{\zeta})H_m(\tilde{\xi})H_p(\tilde{x}_{12})\right], \tag{24}$$

where $\sigma_n=\{(i,j,k,l,m,p)\in\mathbb{N}^6|i+j+k+l+m+p=n\}$ and the GC coefficients are given by $\langle\tilde{x}^i\tilde{x}_1^j\tilde{x}_2^k\tilde{\zeta}^l\tilde{\xi}^m\tilde{x}_{12}^p\rangle_{\rm GC}=\langle H_i(\tilde{x})H_j(\tilde{x}_1)H_k(\tilde{x}_2)H_l(\tilde{\zeta})H_m(\tilde{\xi})H_p(\tilde{x}_{12})\rangle$.

Equations (21) and (24) fully characterize the one-point statistics of a possibly anisotropic weakly NG field in 2D and 3D. These expansions apply whatever the origin of the anisotropy, and in particular for redshift-induced anisotropy in a cosmic environment.

4 PREDICTION FOR MINKOWSKI FUNCTIONALS AND EXTREMA COUNTS

Topological and geometrical measures in redshift space that we are investigating are obtained by integrating the suitable quantities over the distributions (21) and (24), in accordance with equations (1)–(9). We first collect the results for Minkowski functionals, for which such integrals can be carried analytically, and then discuss extrema counts, where one has to resort to numerical integration. The GC expansion leads to a series representation, e.g. $\chi_{\rm 3D}=\chi_{\rm 3D}^{(0)}+\chi_{\rm 3D}^{(1)}+\chi_{\rm 3D}^{(2)}+\cdots$ for which we give here the zero (Gaussian) and the first (NG) order terms. $\chi_{\rm 3D}^{(1)}$ correspond to first-order terms in the variance σ in the cosmological perturbation series. The Gaussian terms, e.g. $\chi_{\rm 3D}^{(0)}$, in redshift space have been first investigated in Matsubara (1996), while the first NG corrections are novel results of this paper. Higher order terms can also be readily obtained within our formalism (see Appendix C). Most of our results are general for arbitrary weakly NG fields with axisymmetric statistical properties.

The presented statistics fall into two families. One group is the statistics of the 3D field as a whole, namely the 3D Euler characteristic, $\chi_{\rm 3D}(\nu)$, the area of the isosurfaces in 3D space, $\mathcal{N}_3(\nu)$, the volume above a threshold, f_V, and the differential count of extrema in 3D. The other group consists of measures on lower dimensional cuts through 3D volumes. These correspond to measures on a planar 2D cut of the 2D Euler characteristic, $\chi_{\rm 2D}(\nu)$, and the length of isocontours on a 2D plane, $\mathcal{N}_2(\nu)$, and statistics of zero crossings, $\mathcal{N}_1(\nu)$, along pencil-beam 1D

[3] 'Field cumulant' means the cumulant computed after expressing the non-linear variable through the field quantities, e.g. $\langle q_\perp^4\rangle_{\rm field\ c}\equiv\langle(x_1^2+x_2^2)^2\rangle_c$. We drop the prefix 'field', always assuming 'field' cumulants.

lines through the volume, as well as the corresponding differential extrema counts. The lower dimensional statistics in an anisotropic space give additional leverage to study anisotropy properties, e.g. in the cosmological context, the magnitude of the redshift distortion, through their dependence on the direction in which the section of the volume is taken. Indeed, we show below [equations (40), (54), ...] that these statistics follow the following generic form (again using $\chi_{2\mathrm{D}}$ as an example and omitting numerical constant factors):

$$\chi_{2\mathrm{D}}(\nu,\theta_{\mathcal{S}}) \propto P(\nu)A(\beta_\sigma,\theta_{\mathcal{S}})\left(H_i(\nu)+\sigma_z\sum_{m=\pm1,\pm3}\left(\tilde{S}^{(m),\perp}_{\chi_{2\mathrm{D}},z}+\tilde{S}^{(m),\parallel}_{\chi_{2\mathrm{D}},z}\right)H_{i+m}(\nu)+\cdots\right), \tag{25}$$

where the overall amplitude $\mathcal{A}$ depends in an angle-sensitive way on the anisotropy parameter,

$$\beta_\sigma \equiv 1-\sigma^2_{1\perp}/2\sigma^2_{1\parallel}, \tag{26}$$

that measures the difference between the rms values of the LOS and perpendicular components of the gradient, and $\theta_{\mathcal{S}}$ denotes the angle between the LOS and the 2D slice under consideration. In the case of isotropy, $\beta_\sigma = 0$. In anisotropic situations, β_σ is positive, spanning the range $0 < \beta_\sigma \leq 1$, when the field changes faster in the z-direction ($\sigma_{1\parallel} > \frac{1}{2}\sigma_{1\perp}$) as is the case, for instance, in the linear regime of redshift corrections where $\beta_\sigma = \frac{4}{5}\beta(1+3\beta/7)/(1+6\beta/5+3\beta^2/7)$, with $\beta = f/b_1$. When the LOS variations are smoother than the perpendicular one ($\sigma_{1\parallel} < \frac{1}{2}\sigma_{1\perp}$), β_σ is negative, $-\infty \leq \beta_\sigma < 0$, as is the case in the non-linear 'finger-of-God' regime.

The Gaussian and subsequent contributions have distinct signatures in this Hermite decomposition. While the Gaussian term is described by an appropriate single Hermite mode $H_i(\nu)$, the first-order corrections excite the modes that have parity opposite to that of the leading Gaussian order. The amplitude of non-Gaussianity is proportional to the variance σ of the field and combinations of third-order cumulants $\tilde{S}^{(m)}_{\chi_{2\mathrm{D}},z}$ that can be different if measured in the directions parallel, $\tilde{S}^{(m),\parallel}_{\chi_{2\mathrm{D}},z}$, or perpendicular, $\tilde{S}^{(m),\perp}_{\chi_{2\mathrm{D}},z}$, to the LOS (e.g. $\langle xq^2_\perp\rangle$ versus $\langle xx^2_3\rangle$).

This suggests the following strategy: (a) determine the β parameter from the amplitude of the 1D or 2D statistics taken at different angles to the LOS; (b) determine the amplitude of non-Gaussianity, and correspondingly σ_z in redshift space, by comparing the Gaussian and NG contributions through fitting distinct Hermite functions to the measurements; and (c) determine the real-space σ by combining the results of the two previous steps. This strategy is implemented on a fiducial experiment in Section 5.5. The 3D statistics have a representation similar to that of equation (25), but without any control over the angle parameter. Therefore, they are less sensitive to the effects of anisotropy, but should provide more robust measurements of the NG corrections than any given lower dimensional subset of data.

4.1 PDF of the field and the filling factor, f_V

The simplest statistic and Minkowski functional is the filling factor, f_V, of the excursion set, i.e. the volume fraction occupied by the region above the threshold ν. Derived from the PDF of the field alone (Bernardeau & Kofman 1995; Juszkiewicz et al. 1995),

$$P(\nu) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\nu^2/2}\left[1+\sum_{i=3}^{\infty}\frac{1}{i!}\left\langle x^i\right\rangle_{\mathrm{GC}}H_i(\nu)\right] = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\nu^2/2}\left[1+\sigma\frac{S_3}{6}H_3(\nu)+\sigma^2\left(\frac{S_4}{24}H_4(\nu)+\frac{S_3^2}{72}H_6(\nu)\right)+\cdots\right], \tag{27}$$

the functional $f_V(\nu) = \int_\nu^\infty P(x)\,\mathrm{d}x$ is given by

$$\begin{aligned} f_V(\nu) &= \frac{1}{2}\mathrm{Erfc}\left(\frac{\nu}{\sqrt{2}}\right)+\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\nu^2/2}\sum_{i=3}^{\infty}\frac{1}{i!}\left\langle x^i\right\rangle_{\mathrm{GC}}H_{i-1}(\nu) \\ &= \frac{1}{2}\mathrm{Erfc}\left(\frac{\nu}{\sqrt{2}}\right)+\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\nu^2/2}\left[\sigma\frac{S_3}{6}H_2(\nu)+\sigma^2\left(\frac{S_4}{24}H_3(\nu)+\frac{S_3^2}{72}H_5(\nu)\right)+\cdots\right], \end{aligned} \tag{28}$$

where the terms from the GC expansion are rearranged in powers of σ (forming the Edgeworth expansion) using the usual skewness $S_3 = \langle x^3\rangle/\sigma$, kurtosis $S_4 = (\langle x^4\rangle_c - 3)/\sigma^2$ and subsequent scaled cumulants of the field x.

There are several advantages of using the value of the filling factor f_V instead of ν as a variable in which to express all other statistics. Indeed, the fraction of volume occupied by a data set is readily available from the data, whereas specifying ν requires prior knowledge of the variance σ, a quantity which is typically the main unknown for such investigations. Following Gott et al. (1987, 1989), Gott (1988), Seto (2000) and Matsubara (2003), let us therefore introduce the threshold $\nu_f = \sqrt{2}\,\mathrm{erfc}^{-1}(2f_V)$ to be used as an observable alternative to ν. The mapping between ν and ν_f is monotonic and is implicitly given by the identity

$$\sqrt{2\pi}\,f_V = \int_{\nu_f}^\infty \mathrm{d}y\,\mathrm{e}^{-y^2/2} = \int_\nu^\infty \mathrm{d}x\,\mathrm{e}^{-x^2/2}+\sum_{i=3}^{\infty}\frac{1}{i!}\left\langle x^i\right\rangle_{\mathrm{GC}}H_{i-1}(\nu)\mathrm{e}^{-\nu^2/2}, \tag{29}$$

which when inverted implies

$$\nu = \nu_f+\sigma\frac{S_3}{6}H_2(\nu_f)+\mathcal{O}(\sigma^2). \tag{30}$$

If truncated to the first NG order, this relation remains monotonic for $\nu_f > -3/\langle x^3\rangle$. All Minkowski functionals derived below will be re-expressed in terms of ν_f in Section 4.5.1.

4.2 Euler characteristic in two and three anisotropic dimensions

4.2.1 3D Euler characteristic: χ_{3D}

From equation (3), the 3D Euler characteristic reads

$$\chi_{3D} = -\frac{1}{\sigma_{1\perp}^2\sigma_{1\parallel}} \int \mathrm{d}x\, \mathrm{d}\zeta\, \mathrm{d}J_{2\perp}\mathrm{d}\xi\, \mathrm{d}Q^2\mathrm{d}\Upsilon\, P_{\rm ext}(x, \zeta, J_{2\perp}\xi, Q^2, \Upsilon) I_3\,, \tag{31}$$

where $I_3 = \det x_{ij}$. In terms of the variables defined in Section 3, the Hessian is

$$I_3 = \frac{1}{4}x_{33}\left(J_{1\perp}^2 - J_{2\perp}\right) - \frac{1}{2}\gamma_2\left(Q^2 J_{1\perp} - \Upsilon\right). \tag{32}$$

The integration is easily performed using the orthogonality properties of Hermite and Laguerre polynomials to give the expression for the 3D Euler characteristic *to all orders in non-Gaussianity* that can be found in Appendix C. The result in equation (C4) has the form

$$\chi_{3D}(\nu) = \frac{1}{2}\mathrm{Erfc}\left(\frac{\nu}{\sqrt{2}}\right)\chi_{3D}(-\infty) + \chi_{3D}^{(0)}(\nu) + \chi_{3D}^{(1)}(\nu) + \cdots\,, \tag{33}$$

where the asymptotic limit $\chi_{3D}(-\infty)$ is zero in the infinite simply connected 3D space, but can reflect the Euler characteristic of the mask if data are available only in subregions with a complex mask.[4]

At Gaussian order, the 3D Euler characteristic reads, in accordance with Matsubara (1996),

$$\chi_{3D}^{(0)}(\nu) = \frac{\sigma_{1\parallel}\sigma_{1\perp}^2}{\sigma^3}\frac{H_2(\nu)}{8\pi^2}\mathrm{e}^{-\nu^2/2}\,. \tag{34}$$

At first NG order, the 3D Euler characteristic is (using relations between cumulants listed in Appendix A2)

$$\chi_{3D}^{(1)}(\nu) = \frac{\sigma_{1\parallel}\sigma_{1\perp}^2}{\sigma^3}\frac{\mathrm{e}^{-\nu^2/2}}{8\pi^2}\left[\frac{H_5(\nu)}{3!}\left\langle x^3\right\rangle + H_3(\nu)\left\langle x\left(q_\perp^2 + x_3^2/2\right)\right\rangle - \frac{H_1(\nu)}{\gamma_\perp}\left\langle J_{1\perp}\left(q_\perp^2 + x_3^2\right)\right\rangle\right], \tag{35}$$

where, when we expand our cumulants in terms of the field variables,

$$\left\langle x\left(q_\perp^2 + x_3^2/2\right)\right\rangle = \frac{\sigma^2}{\sigma_{1\perp}^2}\left\langle x\left(\nabla_\perp x\cdot\nabla_\perp x\right)\right\rangle + \frac{\sigma^2}{2\sigma_{1\parallel}^2}\left\langle x\left(\nabla_\parallel x\cdot\nabla_\parallel x\right)\right\rangle, \tag{36}$$

$$\left\langle J_{1\perp}\left(q_\perp^2 + x_3^2\right)\right\rangle = \frac{\sigma^3}{\sigma_{2\perp}\sigma_{1\perp}^2}\left\langle\Delta_\perp x\ \left(\nabla_\perp x\cdot\nabla_\perp x\right)\right\rangle + \frac{\sigma^3}{\sigma_{2\perp}\sigma_{1\parallel}^2}\left\langle\Delta_\perp x\ \left(\nabla_\parallel x\cdot\nabla_\parallel x\right)\right\rangle. \tag{37}$$

Note that the $n = 3$ GC coefficients are actually equal to the cumulants of field variables; therefore, the brackets without the label GC are used, meaning these are standard cumulants. This correspondence is preserved at the next order as well, but will eventually be broken for the higher order GC coefficients. We refer again to Gay et al. (2012) for a detailed discussion.

The isotropic limit can be put in a concise form using equation (16) and the relationships $\langle\Delta_\perp x(\nabla_\parallel x\cdot\nabla_\parallel x)\rangle = \langle\Delta_\perp x(\nabla_\perp x\cdot\nabla_\perp x)\rangle = \langle(\nabla x\cdot\nabla x)\Delta x\rangle/3$:

$$\chi_{3D}^{\rm iso}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^2}\frac{\sigma_1^3}{3\sqrt{3}\sigma^3}\left[H_2(\nu) + \frac{1}{3!}H_5(\nu)\left\langle x^3\right\rangle + \frac{3}{2}\frac{\sigma^2}{\sigma_1^2}H_3(\nu)\left\langle x\left(\nabla x\cdot\nabla x\right)\right\rangle - \frac{9}{4}\frac{\sigma^4}{\sigma_1^4}H_1(\nu)\left\langle\left(\nabla x\cdot\nabla x\right)\Delta x\right\rangle + \mathcal{O}(\sigma^2)\right], \tag{38}$$

which is in exact agreement with Gay et al. (2012) (see Appendix A2).

These predictions for the 3D Euler characteristics are now first validated on toy models for which simulations are straightforward and cumulants simply analytic (see Appendix B for details about this $f_{\rm nl}$ toy model). In this non-dynamical model, the redshift correction is simulated by transforming the density according to the linear equation (11) with an f factor chosen freely. Figs B1 and B2 present a good match between the theoretical predictions of equations (35) and (38) for the 3D NG anisotropic Euler characteristic to simulated fields. They also show the evolution as a function of $f_{\rm nl}$ (i.e. as a function of the non-Gaussianity) for two values of the anisotropy parameter, $f = 0$ (real space) and $f = 1$ (redshift space).

Fig. 4 summarizes comparison of many results of this section with measurement on the density fields in DM-only, 256^3, scale-invariant LSS simulations with power index $n = -1$ and $\Omega_{\rm m} = 1$. To reproduce the redshift effects, the positions of the particles are shifted along z according to equation (10). For DM, we have $f = \beta = 1$. The lowest-right panel in Fig. 4 compares the prediction of equation (35) to the χ_{3D} statistics. The measurements are done at epochs for which $\sigma = 0.18$ in real space. For these fields, non-Gaussianity comes from gravitational clustering but also from the mapping into redshift space which is intrinsically non-linear. The theoretical formula uses the cumulants measured from simulation itself, so we do not test how accurately the cumulants can be predicted, e.g. using PT. Fig. 4 shows that the theoretical prediction at next-to-leading order (NLO; i.e. at first NG order) mimics very well the measurement for intermediate contrasts

[4] Here by 'masking' we understand the procedure that excludes some regions of space from observations without modifying the underlying statistical properties of the field.

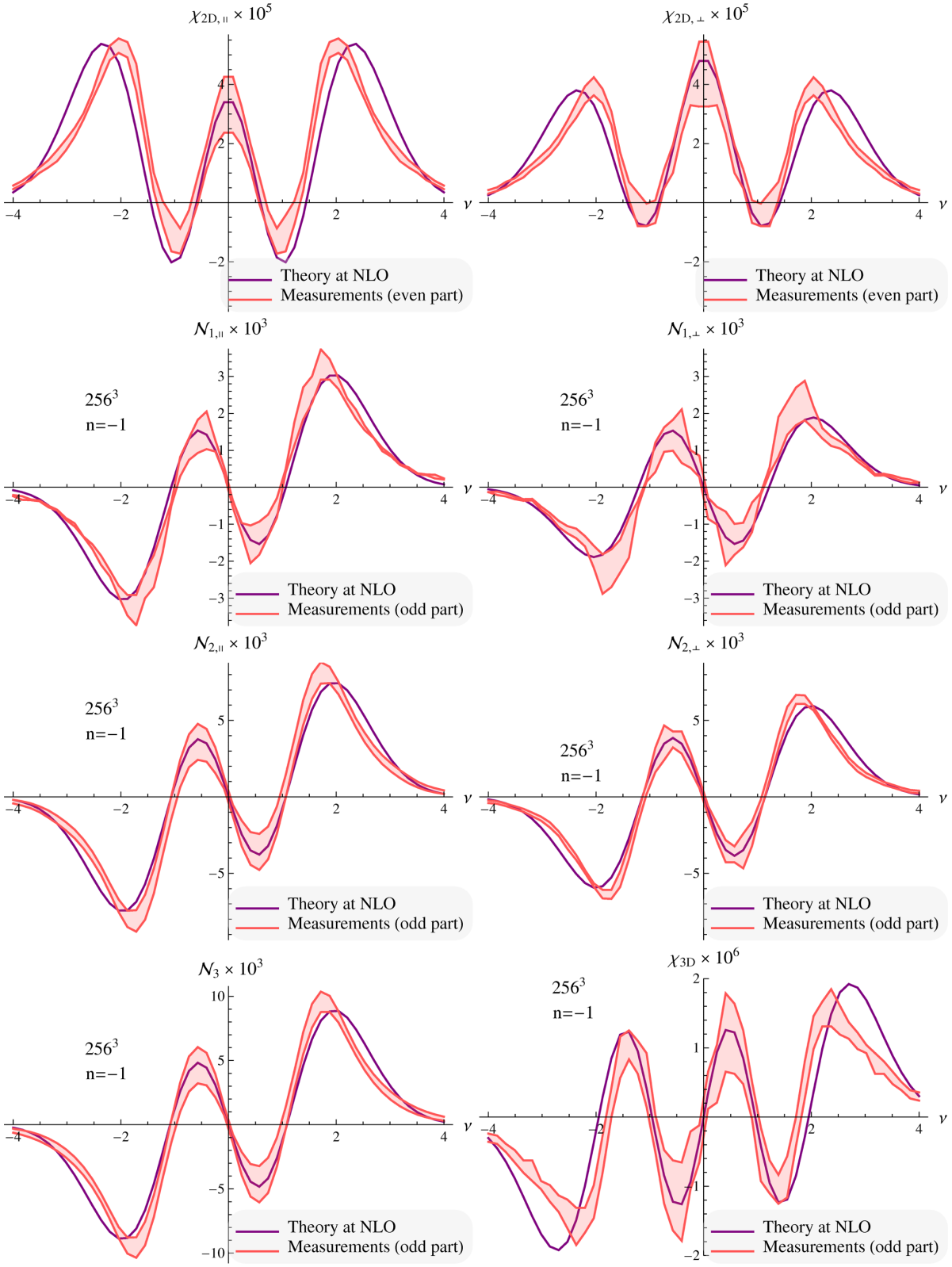


Figure 4. Gravity-induced NLO NG corrections for the Minkowski functionals of a mildly non-linear 3D density field ($n = -1$, $\sigma = 0.18$) in redshift space ($f = 1$). Theoretical predictions are displayed with solid purple lines and measurements with shaded one-sigma dispersion. Top left: 2D Euler characteristic in planes parallel to the LOS; top right: 2D Euler characteristic in planes perpendicular to the LOS; central top left: $\mathcal{N}_1$ (contour crossing) in planes parallel to the LOS; central top right: $\mathcal{N}_1$ in planes perpendicular to the LOS; central bottom left: $\mathcal{N}_2$ (length of isocontour) in planes parallel to the LOS; central bottom right: $\mathcal{N}_2$ in planes perpendicular to the LOS; bottom left: $\mathcal{N}_3$ (area of isocontour); bottom right: 3D Euler characteristic. Note that the predicted first-order correction fits very well the odd part (respectively even for the 2D Euler characteristic) of the measured correction for intermediate contrast ($-1.5 \lesssim \nu \lesssim 1.5$).

as expected (note that we plot the odd part of the signal to suppress the Gaussian term and focus on deviation to Gaussianity only). To better fit the data at the tails of the distribution, one has to take into account higher order corrections, which, as previous real-space studies show, are non-negligible for $\sigma \geq 0.18$.

4.2.2 2D Euler characteristic: $\chi_{2\mathrm{D}}$

Let us now investigate the Euler characteristic of the field on a 2D planar section, $\mathcal{S}$, of a 3D z-anisotropic space. As redshift distortion occurs along the LOS, this statistics thus depends on the angle $\theta_{\mathcal{S}}$ between the plane $\mathcal{S}$ and that LOS. The 2D field restricted to the plane $\mathcal{S}$ is denoted as $\tilde{x}(s_1, s_2)$, where s_1 and s_2 are two Cartesian coordinates on the plane, chosen so that s_1 spans the direction perpendicular to the LOS. From equation (5), the Euler characteristic on the plane ($\mathcal{S}$) reads

$$\chi_{2\mathrm{D}} = \frac{1}{\tilde{\sigma}_{1\perp}\tilde{\sigma}_{1\parallel}} \int \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{x}_1\, \mathrm{d}\tilde{x}_2\, \mathrm{d}\tilde{\zeta}\, \mathrm{d}\tilde{\xi}\, \mathrm{d}\tilde{x}_{12} \tilde{I}_2 \delta_{\mathrm{D}}(\tilde{x}_1)\delta_{\mathrm{D}}(\tilde{x}_2) P(\tilde{x}, \tilde{x}_1, \tilde{x}_2, \tilde{\zeta}, \tilde{\xi}, \tilde{x}_{12}), \tag{39}$$

where $\tilde{I}_2 = \tilde{\sigma}_{2\perp}\tilde{\sigma}_{2\parallel}(\tilde{x}_{11}\tilde{x}_{22} - \tilde{\gamma}_2\tilde{x}_{12}^2)$ is the Hessian determinant on the plane. Rewriting $\tilde{I}_2$ in terms of $\tilde{x}, \tilde{\zeta}, \tilde{\xi}$ and $\tilde{x}_{12}$ and using the orthogonality of Hermite polynomials, one obtains, after some algebra, an all-order expansion,

$$\begin{aligned}
\chi_{2\mathrm{D}}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}} \frac{\tilde{\sigma}_{2\perp}\tilde{\sigma}_{2\parallel}}{\tilde{\sigma}_{1\perp}\tilde{\sigma}_{1\parallel}} \times \Bigg(& \tilde{\gamma}_\parallel \tilde{\gamma}_\perp H_1(\nu) + \tilde{\gamma}_\parallel \tilde{\gamma}_\perp \sum_{n=3}^{\infty} \sum_{\sigma_n} \frac{(-1)^{k+j}}{i!\; j!\; k!\; 2^k 2^j} \left\langle \tilde{x}^i \tilde{x}_1^{2j} \tilde{x}_2^{2k} \right\rangle_{\mathrm{GC}} (H_{i+1}(\nu) + 2i H_{i-1}(\nu) + i(i-1) H_{i-3}(\nu)) \\
& - \sum_{n=3}^{\infty} \sum_{\sigma_{n-1}} \frac{(-1)^{k+j} (H_i(\nu) + i H_{i-2}(\nu))}{i!\; j!\; k!\; 2^k 2^j \sqrt{1-\tilde{\gamma}_\perp^2}} \Big[\tilde{\gamma}_\perp \sqrt{1 - \tilde{\gamma}_\perp^2 - \tilde{\gamma}_\parallel^2 - \tilde{\gamma}_2^2 + 2\tilde{\gamma}_\perp \tilde{\gamma}_\parallel \tilde{\gamma}_2} \left\langle \tilde{x}^i \tilde{x}_1^{2j} \tilde{x}_2^{2k} \tilde{\xi} \right\rangle_{\mathrm{GC}} + \left(\tilde{\gamma}_\perp (\tilde{\gamma}_2 - \tilde{\gamma}_\parallel \tilde{\gamma}_\perp) + \tilde{\gamma}_\parallel (1 - \tilde{\gamma}_\perp^2) \right) . \\
& \times \left\langle x^i \tilde{x}_1^{2j} \tilde{x}_2^{2k} \tilde{\zeta} \right\rangle_{\mathrm{GC}} \Big] + \sum_{n=3}^{\infty} \sum_{\sigma_{n-2}} \frac{(-1)^{k+j}}{i!\; j!\; k!\; 2^k 2^j} H_{i-1}(\nu) \Big[(\tilde{\gamma}_2 - \tilde{\gamma}_\parallel \tilde{\gamma}_\perp) \left\langle \tilde{x}^i \tilde{x}_1^{2j} \tilde{x}_2^{2k} \tilde{\zeta}^2 \right\rangle_{\mathrm{GC}} - \tilde{\gamma}_2 \left\langle \tilde{x}^i \tilde{x}_1^{2j} \tilde{x}_2^{2k} \tilde{x}_{12}^2 \right\rangle_{\mathrm{GC}} \\
& + \sqrt{1 - \tilde{\gamma}_\perp^2 - \tilde{\gamma}_\parallel^2 - \tilde{\gamma}_2^2 + 2\tilde{\gamma}_\perp \tilde{\gamma}_\parallel \tilde{\gamma}_2} \left\langle \tilde{x}^i \tilde{x}_1^{2j} \tilde{x}_2^{2k} \tilde{\zeta} \tilde{\xi} \right\rangle_{\mathrm{GC}} \Big] \Bigg),
\end{aligned} \tag{40}$$

where $\sigma_n = \{(i, j, k) \in \mathbb{N}^3 | i + 2j + 2k = n\}$ and $H_{-1}(\nu) = \sqrt{\pi/2}\ \mathrm{Erfc}(\nu/\sqrt{2})$. In the Gaussian limit, the Euler characteristic of a 2D plane then reads

$$\chi_{2\mathrm{D}}^{(0)}(\nu, \theta_{\mathcal{S}}) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}} \frac{\tilde{\sigma}_{1\perp}\tilde{\sigma}_{1\parallel}}{\sigma^2} H_1(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}} \frac{\sigma_{1\perp}\sigma_{1\parallel}}{\sqrt{2}\sigma^2} \sqrt{1 - \beta_\sigma \sin^2\theta_{\mathcal{S}}}\ H_1(\nu), \tag{41}$$

where β_σ is defined in equation (26). This is in agreement with Matsubara (1996). Note that the amplitude of this Gaussian term is overestimated when assuming isotropy.

The $n = 3$ term (i.e. the first correction from Gaussianity) in the expansion gives, in 'on-plane' variables,

$$\chi_{2\mathrm{D}}^{(1)}(\nu, \theta_{\mathcal{S}}) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}} \frac{\tilde{\sigma}_{1\perp}\tilde{\sigma}_{1\parallel}}{\sigma^2} \times \left[\frac{1}{3!} H_4(\nu) \left\langle \tilde{x}^3 \right\rangle + \frac{1}{2} H_2(\nu) \left\langle \tilde{x} \left(\tilde{x}_1^2 + \tilde{x}_2^2 \right) \right\rangle - \frac{1}{2} \tilde{\gamma}_\perp^{-1} \left\langle \tilde{x}_{11} \tilde{x}_2^2 \right\rangle \right], \tag{42}$$

or, using the cumulants of the 3D field,

$$\begin{aligned}
\chi_{2\mathrm{D}}^{(1)}(\nu, \theta_{\mathcal{S}}) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}} \frac{\sigma_{1\perp}\sigma_{1\parallel}}{\sqrt{2}\sigma^2} \sqrt{1 - \beta_\sigma \sin^2\theta_{\mathcal{S}}} \times \Bigg[& \frac{1}{3!} H_4(\nu) \left\langle x^3 \right\rangle + H_2(\nu) \left(\left\langle x q_\perp^2 \right\rangle + \frac{1}{2} \frac{\cos^2\theta_{\mathcal{S}}}{1 - \beta_\sigma \sin^2\theta_{\mathcal{S}}} \left(\left\langle x x_3^2 \right\rangle - \left\langle x q_\perp^2 \right\rangle \right) \right) \\
& - \frac{1}{\gamma_\perp} \left(\left\langle q_\perp^2 J_{1\perp} \right\rangle + \frac{1}{2} \frac{\cos^2\theta_{\mathcal{S}}}{1 - \beta_\sigma \sin^2\theta_{\mathcal{S}}} \left(\left\langle x_3^2 J_{1\perp} \right\rangle - 2 \left\langle q_\perp^2 J_{1\perp} \right\rangle \right) \right) \Bigg] \qquad ,
\end{aligned} \tag{43}$$

where $\sigma_{1\perp}^2 q_\perp^2 = \sigma^2(\nabla_\perp x \cdot \nabla_\perp x)$, $\sigma_{1\parallel}^2 x_3^2 = \sigma^2(\nabla_\parallel x \cdot \nabla_\parallel x)$ and $\sigma_{2\perp} J_{1\perp} = \sigma\ \Delta x$. For the particular case when the cut is done through an isotropic 3D field, the 2D Euler characteristic to first order in non-Gaussianity is

$$\chi_{2\mathrm{D}}^{\mathrm{iso}}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}} \frac{\sigma_1^2}{3\sigma^2} \left[H_1(\nu) + \frac{1}{3!} H_4(\nu) \langle x^3 \rangle + \frac{\sigma^2}{\sigma_1^2} H_2(\nu) \langle x\, (\nabla x \cdot \nabla x) \rangle - \frac{3\sigma^4}{4\sigma_1^4} \langle \Delta x\, (\nabla x \cdot \nabla x) \rangle + \mathcal{O}(\sigma^2) \right], \tag{44}$$

which is in agreement with Gay et al. (2012) (again relying on Appendix A2 for some relations between the cumulants). Figs B1 and B2 show the dependence of the 2D Euler characteristic on f_{nl} and f following equations (43) and (44). As expected, the effect of redshift distortion is enhanced in 2D relative to 3D. This is expected since in 3D, two dimensions remain isotropic. Fig. B3 also shows the evolution of the 2D Euler characteristic as a function of g_{nl} and f. It demonstrates that this set of simulations is well fitted by the sum of the two contributions from f_{nl} and g_{nl}.

For cosmological fields, Fig. 4 compares the prediction of equation (43) for the $\chi_{2\mathrm{D}}$ statistics to scale-invariant LSS simulations in planes parallel and perpendicular to the LOS. Again, this figure shows very good agreement between measurements and predictions at NLO for intermediate contrasts ($-1.5 \lesssim \nu \lesssim 1.5$). There is a noticeable angle dependence of the 2D Euler characteristic which suggests that the

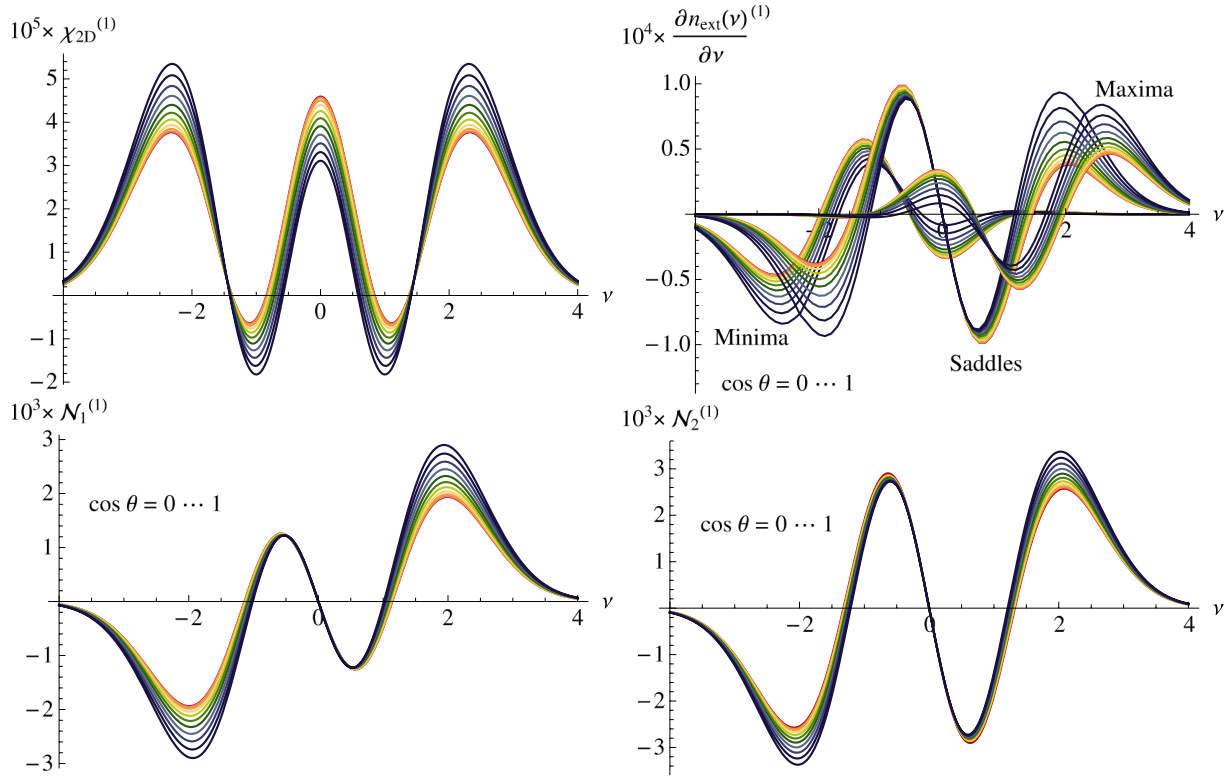


Figure 5. 2D Euler characteristic (top-left panel), 2D extrema counts (top-right panel), length of 2D isocontour (bottom-left panel) and contour crossing (bottom-right panel) at first NG order as a function of $\cos\theta_{\mathcal{S}}$ from 0 (red) to 1 (dark blue). Cumulants used in equations (43), (54) and (60) are measured in a scale-invariant LSS simulation with $\sigma = 0.18$ The prediction for 2D extrema statistics was obtained by numerical integration following Section 4.4. The noticeable dependence of these functionals on the angle $\theta_{\mathcal{S}}$ gives us a mean to measure $\beta = f/b_1$ as suggested in Section 5.5.

procedure described in the introduction of Section 4 is feasible. This angle dependence is also illustrated in Fig. 5, where one can see how the prediction at NLO varies with the angle $\theta_{\mathcal{S}}$ between the LOS and the 2D slice under consideration.

4.3 Other Minkowski functionals

4.3.1 Area of isodensity contours in 3D: $\mathcal{N}_3$

One of the Minkowski functionals is $\mathcal{N}_3(\nu)$, the area (per unit volume) of a 3D isosurface of the density field at level ν. To compute this functional, it is sufficient to consider the JPDF of the field and its first derivatives,

$$P\left(x, q_\perp^2, x_3\right) = \frac{1}{2\pi}\exp\left(-\frac{x^2}{2} - q_\perp^2 - \frac{x_3^2}{2}\right)\left[1 + \sum_{n=3}^{\infty}\sum_{\sigma_n}\frac{(-1)^j}{i!\,j!\,k!}\left\langle x^i q_\perp^{2j} x_3^k\right\rangle_{\rm GC} H_i(x) L_j\left(q_\perp^2\right) H_k(x_3)\right]. \tag{45}$$

From equation (2), the area of a 3D isosurface is

$$\mathcal{N}_3(\nu) = \frac{1}{\sigma}\int \mathrm{d}q_\perp^2 \mathrm{d}x_3 P(\nu, q_\perp^2, x_3)\sqrt{\sigma_{1\perp}^2 q_\perp^2 + \sigma_{1\parallel}^2 x_3^2}\,. \tag{46}$$

Computing the integrals, we express the results using the anisotropy parameter $\beta_\sigma = 1 - \sigma_{1\perp}^2/2\sigma_{1\parallel}^2$. At Gaussian order, expression (46) yields the result, consistent with Matsubara (1996),

$$\mathcal{N}_3^{(0)}(\nu) = \frac{2}{\pi}\frac{\sigma_1}{\sqrt{3}\sigma}\frac{1 - A(\beta_\sigma)}{\sqrt{1 - 2\beta_\sigma/3}}\mathrm{e}^{-\nu^2/2}, \quad \text{with} \quad A(\beta_\sigma) \equiv \frac{1}{2}(\beta_\sigma - T(\beta_\sigma) + \beta_\sigma T(\beta_\sigma)), \tag{47}$$

where the function T is defined as $T(\beta_\sigma) = 1/\sqrt{\beta_\sigma}\tanh^{-1}(\sqrt{\beta_\sigma}) - 1$ for $\beta_\sigma \geq 0$ and $T(\beta_\sigma) = 1/\sqrt{|\beta_\sigma|}\tan^{-1}(\sqrt{|\beta_\sigma|}) - 1$ for $\beta_\sigma < 0$. Under this definition, $A(\beta_\sigma)$ describes an $\sim \beta_\sigma/3 + \beta_\sigma^2/15 + \cdots$ correction at small anisotropy $\beta_\sigma \to 0$. We see that in the Gaussian limit, the anisotropy has a very little effect on $\mathcal{N}_3$. The amplitude deviates from unity by less than 1 per cent in the range $-1 < \beta_\sigma < 0.5$, as its series expansion $\propto 1 - \beta_\sigma^2/90, \ldots$ attests. Even at extreme anisotropies, it changes just to $\approx$0.92 at $\beta_\sigma \to -\infty$ and $\approx$0.87 at $\beta_\sigma = 1$.

The computation of the $n = 3$ term corresponding to the first NG correction is also straightforward,

$$\mathcal{N}_3^{(1)}(\nu) = \mathcal{N}_3^{(0)}(\nu)\left[\frac{1}{3!}\langle x^3\rangle H_3(\nu) + \frac{1}{2}\left(\langle x q_\perp^2\rangle + \frac{A(\beta_\sigma)/\beta_\sigma}{1 - A(\beta_\sigma)}\left(\langle x x_3^2\rangle - \langle x q_\perp^2\rangle\right)\right) H_1(\nu)\right]. \tag{48}$$

To first order in β_σ (the anisotropy parameter), one gets the following explicit expression:

$$\mathcal{N}_3^{(1)}(\nu) = \frac{2\mathrm{e}^{-\nu^2/2}\sigma_1}{\sqrt{3}\pi\sigma}\left[1 + \frac{1}{3!}\left\langle x^3\right\rangle H_3(\nu) + \frac{1}{2}\left(\frac{2}{3}\left\langle xq_\perp^2\right\rangle\left[1 - \frac{4}{15}\beta_\sigma\right] + \frac{1}{3}\left\langle xx_3^2\right\rangle\left[1 + \frac{8}{15}\beta_\sigma\right]\right)H_1(\nu)\right] + \mathcal{O}\left(\beta_\sigma^2\right), \tag{49}$$

from which the isotropic limit of Gay et al. (2012) is readily recovered by setting $\beta_\sigma = 0$ and $\langle xq_\perp^2\rangle = \langle xx_3^2\rangle = \langle xq^2\rangle$. We note that anisotropy effects in the $\mathcal{N}_3$ statistics are almost exclusively concentrated in the gradient terms $\propto H_1(\nu)$. This suggests, for example, that the recovery of the skewness $\langle x^3\rangle$ by fitting the $H_3(\nu)$ mode to $\mathcal{N}_3$ will be practically unaffected by redshift distortions. In contrast, one must focus on the $H_1(\nu)$ mode to measure anisotropic effects.

Fig. 4 also compares the prediction of equation (49) for the $\mathcal{N}_3$ statistics to scale-invariant LSS simulations (256^3, $n = -1$). Once again, in this mildly non-linear regime ($\sigma = 0.18$ in real space), the prediction at NLO matches very well the measurement for intermediate contrasts ($-2 \lesssim \nu \lesssim 2$).

4.3.2 Length of isodensity contours in 2D planes: $\mathcal{N}_2$

Let us consider the length (per unit volume) of isodensity contours in 2D slices of the density field, $\mathcal{N}_2$. This functional is the 2D version of the Minkowski functional $\mathcal{N}_3$ for the 3D field. Here the 2D slice is defined by the angle, $\theta_\mathcal{S}$, it makes with the z-axis, and the statistical properties of the field depend on this angle. Again, let us start with the JPDF of the field and its gradient on a 2D plane. Using the same variables as in Section 4.2.2,

$$P(\tilde{x}, \tilde{x}_1, \tilde{x}_2) = \frac{1}{(2\pi)^{3/2}}\exp\left(-\frac{\tilde{x}^2}{2} - \frac{\tilde{x}_1^2}{2} - \frac{\tilde{x}_2^2}{2}\right)\left[1 + \sum_{n=3}^{\infty}\sum_{\sigma_n}\frac{1}{i!\,j!\,k!}\left\langle \tilde{x}^i\tilde{x}_1^j\tilde{x}_2^k\right\rangle_{\mathrm{GC}} H_i(\tilde{x})H_j(\tilde{x}_1)H_k(\tilde{x}_2)\right], \tag{50}$$

where $\sigma_n = \{(i, j, k) \in \mathbb{N}^3 | i + 2j + k = n\}$. From equation (2), the length of isodensity contours in 2D planes is now

$$\mathcal{N}_2(\nu, \theta_\mathcal{S}) = \frac{1}{\sigma}\int \mathrm{d}\tilde{x}_1\,\mathrm{d}\tilde{x}_2 P(\nu, \tilde{x}_1, \tilde{x}_2)\sqrt{\tilde{\sigma}_{1\perp}^2\tilde{x}_1^2 + \tilde{\sigma}_{1\parallel}^2\tilde{x}_2^2}\,. \tag{51}$$

We shall proceed similarly to the 3D case, defining the 2D anisotropy parameter $\tilde{\beta}_\sigma(\theta_\mathcal{S})$ which depends on the orientation of the plane,

$$\tilde{\beta}_\sigma(\theta_\mathcal{S}) \equiv 1 - \frac{\tilde{\sigma}_{1\perp}^2}{\tilde{\sigma}_{1\parallel}^2} = \frac{\beta_\sigma\cos^2\theta_\mathcal{S}}{1 - \beta_\sigma\sin^2\theta_\mathcal{S}}\,. \tag{52}$$

At Gaussian order, the evaluation of equation (51) yields, in agreement with Matsubara (1996),

$$\mathcal{N}_2^{(0)}(\nu, \theta_\mathcal{S}) = \frac{\tilde{\sigma}_{1\parallel}}{\pi\sigma}\mathrm{e}^{-\nu^2/2}E\left(\tilde{\beta}_\sigma\right) = \frac{\sigma_{1\perp}}{\sqrt{2}\pi\sigma}\mathrm{e}^{-\nu^2/2}\frac{E\left(\tilde{\beta}_\sigma\right)}{\sqrt{1 - \tilde{\beta}_\sigma}}, \tag{53}$$

where E is the complete elliptic integral of the second kind. The amplitude behaves as $E(\tilde{\beta}_\sigma)/\sqrt{1 - \tilde{\beta}_\sigma} \sim \pi/2 + \pi\tilde{\beta}_\sigma/8$ at small $\tilde{\beta}_\sigma$ and is thus strongly dependent on the anisotropy parameter. This is distinct from the behaviour of $\mathcal{N}_3$.

Then, the first NG correction (corresponding to $n = 3$) reads

$$\mathcal{N}_2^{(1)}(\nu, \theta_\mathcal{S}) = \mathcal{N}_2^{(0)}\left[\frac{H_3(\nu)}{3!}\left\langle\tilde{x}^3\right\rangle + \frac{1}{2}H_1(\nu)\left(\left\langle\tilde{x}\tilde{x}_2^2\right\rangle + B(\tilde{\beta}_\sigma)\left(\left\langle\tilde{x}\tilde{x}_1^2\right\rangle - \left\langle x\tilde{x}_2^2\right\rangle\right)\right)\right], \text{ with } B(\tilde{\beta}_\sigma) = 2(\tilde{\beta}_\sigma - 1)\frac{\mathrm{d}\ln E(\tilde{\beta}_\sigma)}{\mathrm{d}\tilde{\beta}_\sigma}\,. \tag{54}$$

Here on-plane cumulants are related to 3D ones by $\langle\tilde{x}\tilde{x}_1^2\rangle = \langle xq_\perp^2\rangle$ and $\langle\tilde{x}\tilde{x}_2^2\rangle = (1 - \tilde{\beta}_\sigma/\beta_\sigma)\langle xq_\perp^2\rangle + \tilde{\beta}_\sigma/\beta_\sigma\langle xx_3^2\rangle$. In the small anisotropy limit, equation (54) becomes

$$\mathcal{N}_2^{(1)}(\nu, \theta_\mathcal{S}) = \frac{\sigma_{1\perp}}{2\sqrt{2}\sigma}\mathrm{e}^{-\nu^2/2}\left(1 + \frac{1}{4}\tilde{\beta}_\sigma\right)\left[\frac{H_3(\nu)}{3!}\left\langle\tilde{x}^3\right\rangle + \frac{1}{4}H_1(\nu)\left(\left\langle\tilde{x}\tilde{x}_1^2\right\rangle + \left\langle\tilde{x}\tilde{x}_2^2\right\rangle - \frac{3}{8}\tilde{\beta}_\sigma\left(\left\langle\tilde{x}\tilde{x}_1^2\right\rangle - \left\langle\tilde{x}\tilde{x}_2^2\right\rangle\right)\right)\right] + \mathcal{O}\left(\tilde{\beta}_\sigma^2\right), \tag{55}$$

where the isotropic limit of Gay et al. (2012) is readily recognized at $\tilde{\beta}_\sigma = 0$, $\langle\tilde{x}\tilde{x}_1^2\rangle = \langle\tilde{x}\tilde{x}_2^2\rangle = \langle xq^2\rangle$ and $\sigma_{1\perp} = \sqrt{2/3}\sigma_1$. We should stress that anisotropic effects in equation (55) are contained not only in the $\tilde{\beta}_\sigma$ factors, but also in the deviation of cumulants from their isotropic values. Note that, with the 3D variables, equation (55) becomes

$$\begin{aligned}\mathcal{N}_2^{(1)}(\nu, \theta_\mathcal{S}) = \frac{\sigma_{1\perp}}{2\sqrt{2}\sigma}\mathrm{e}^{-\nu^2/2}\left(1 + \beta_\sigma\frac{\cos^2\theta_\mathcal{S}}{4}\right)\Bigg[&\frac{H_3(\nu)}{3!}\left\langle x^3\right\rangle + \frac{1}{4}H_1(\nu)\left(\left\langle xq_\perp^2\right\rangle + \left\langle xx_3^2\right\rangle\right.\\ &\left.+\left(\sin^2\theta_\mathcal{S} - \beta_\sigma\frac{\cos^2\theta_\mathcal{S}(3 + 5\sin^2\theta_\mathcal{S})}{8}\right)\left(\left\langle xq_\perp^2\right\rangle - \left\langle xx_3^2\right\rangle\right)\right)\Bigg] + \mathcal{O}\left(\beta_\sigma^2\right).\end{aligned} \tag{56}$$

The most interesting case for the $\mathcal{N}_2$ statistics is when the slice is passing through the observer, and, correspondingly, contains the LOS. This is the case for 2D slices in observational catalogues such as the Sloan Digital Sky Survey. This setup corresponds to $\theta_\mathcal{S} = 0$ and $\tilde{\beta}_\sigma = \beta_\sigma$. Indeed, the measurement of $\mathcal{N}_2$ in such slices gives direct access to the 3D anisotropy parameter β_σ and, by extension, $\Omega_\mathrm{m}^{0.6}/b$, even when full

3D data are not available. The study of 2D slices at different angles θ_S is possible if the 3D cube of data is available and offers an alternative way of analysing such cubes. Varying θ_S allows one to introduce functional dependence of statistics on the parameter, which gives additional access to $\tilde{\beta}_\sigma$ through variation of the amplitude of the statistics even when the normalization scale $\sigma_{1\perp}/\sigma$ is poorly determined.

Fig. 4 compares the prediction of equation (54) for the $\mathcal{N}_2$ statistics to scale-invariant LSS simulations in planes parallel and perpendicular to the LOS. In both cases, measurements are well fitted by the prediction at the first NG order for intermediate contrasts. One can easily notice that this statistics varies macroscopically with the angle θ_S. This dependence is also shown in Fig. 5 for the NLO prediction. Indeed, Fig. 5 displays the variation of the NG contribution to the $\mathcal{N}_2$ statistics as a function of the orientation of the 2D slices relative to the LOS. Note that anisotropy affects the $H_1(\nu)$ harmonic of this dependence and can be detected by a linear fit to $\mathcal{N}_2^{(1)}(\nu, \theta_S)$ at different θ_S.

4.3.3 Contour crossings: $\mathcal{N}_1$

Contour-crossing statistics measures the average number of times a given line crosses the isocontours of a field, $\mathcal{N}_1$. It is closely related to the average area of the isocontours per unit volume, $\mathcal{N}_3$. It is equivalent to $\mathcal{N}_3$ when averaged over all possible orientations of the chosen line (indeed, the surface area per unit volume $\mathcal{N}_3$ can be understood as an average 'hyperflux' of isosurfaces, i.e. how many times per unit length they cross a random line), but has distinct dependence on line orientation if the field is anisotropic. A special advantage of $\mathcal{N}_1$ is that it can be applied when only 'pencil-beam' data are available, as for example in Lyα lines, although in this case one is limited to use only the LOS direction of the field. In general, the behaviour of $\mathcal{N}_1$ along lines with arbitrary orientation relative to the LOS contains additional information.

To compute contour-crossing statistics, we assume (without loss of generality) that the line intersecting the field lies in the $(\boldsymbol{u}_1, \boldsymbol{u}_3)$ plane with a direction defined by the unit vector $\hat{l} = (\sin\theta_{\mathcal{L}}, 0, \cos\theta_{\mathcal{L}})$. Then the mean number of intersections between a line ($\mathcal{L}$) and the isodensity contours is simply the average of the field gradient projection $|\nabla x \cdot \hat{l}| = |\sigma_{1\perp} x_1 \sin\theta_L/\sqrt{2} + \sigma_{1\parallel} x_3 \cos\theta_L|$. In contrast to the 2D statistics case, we shall not introduce tilde variables on the line, and write the statistics immediately in terms of the 3D variables. Starting from the JPDF of the field and two components of its first derivatives

$$P(x, x_1, x_3) = \frac{1}{(2\pi)^{3/2}} \exp\left(-\frac{x^2}{2} - \frac{x_1^2}{2} - \frac{x_3^2}{2}\right) \left[1 + \sum_{n=3}^{\infty} \sum_{i,j,k=0}^{i+j+k=n} \frac{1}{i!\,j!\,k!} \left\langle x^i x_1^j x_3^k \right\rangle_{\mathrm{GC}} H_i(x) H_j(x_1) H_k(x_3)\right], \tag{57}$$

from equation (2), we write

$$\mathcal{N}_1(\nu, \theta_{\mathcal{L}}) = \frac{1}{\sigma} \int \mathrm{d}x_1\, \mathrm{d}x_3 P(\nu, x_1, x_3) \left|\frac{\sigma_{1\perp}}{\sqrt{2}} x_1 \sin\theta_{\mathcal{L}} + \sigma_{1\parallel} x_3 \cos\theta_{\mathcal{L}}\right|. \tag{58}$$

At Gaussian order, using again $\beta_\sigma = 1 - \sigma_{1\perp}^2/2\sigma_{1\parallel}^2$, we find, in agreement with Matsubara (1996),

$$\mathcal{N}_1^{(0)}(\nu, \theta_{\mathcal{L}}) = \frac{\sigma_{1\parallel}}{\pi\sigma} \mathrm{e}^{-\nu^2/2} \sqrt{1 - \beta_\sigma \sin^2\theta_L}. \tag{59}$$

Focusing on the next order term ($n = 3$) and using the relationships $\langle x x_1 x_3\rangle = \langle x^2 x_i\rangle = 0$, one gets

$$\mathcal{N}_1^{(1)}(\nu, \theta_{\mathcal{L}}) = \mathcal{N}_1^{(0)}(\theta_{\mathcal{L}}, \nu) \left[\frac{1}{3!}\left\langle x^3\right\rangle H_3(\nu) + \frac{1}{2} H_1(\nu) \left(\left\langle x q_\perp^2\right\rangle + \frac{\cos^2\theta_{\mathcal{L}}}{1 - \beta_\sigma \sin^2\theta_{\mathcal{L}}} \left(\left\langle x x_3^2\right\rangle - \left\langle x q_\perp^2\right\rangle\right)\right)\right]. \tag{60}$$

In the isotropic limit where $\beta_\sigma = 0$ and $\langle x q_\perp^2\rangle = \langle x x_3^2\rangle = \langle x q^2\rangle$, equations (59) and (60) together becomes

$$\mathcal{N}_1^{\mathrm{iso}}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}\sigma_1}{\sqrt{3}\pi\sigma} \left[1 + \frac{1}{3!} H_3(\nu) \left\langle x^3\right\rangle + \frac{1}{2} H_1(\nu) \left\langle x q^2\right\rangle + \mathcal{O}(\sigma^2)\right]. \tag{61}$$

Fig. 4 compares the prediction of equation (60) for the $\mathcal{N}_1$ statistics to scale-invariant LSS simulations in planes parallel and perpendicular to the LOS. The agreement between the prediction at NLO and measurements is very good for contrasts in the range $-2 \lesssim \nu \lesssim 2$. The tails of the distribution are more sensitive to higher order terms in the GC expansion. Planes parallel and perpendicular to the LOS give rise to a noticeable difference in the first-order correction for the contour-crossing statistics. Fig. 5 also displays the variation of $\mathcal{N}_1$ as a function of the orientation of the 2D slices relative to the LOS.

4.4 Extrema counts

Extrema counts are given by averaging the absolute value of the determinant of the Hessian (I_3 and I_2 in 3D and 2D, respectively) under the condition of zero gradient over the range of Hessian eigenvalues space that maintains the correspondent signature of the sorted eigenvalues, namely, in 3D, $\lambda_1 > \lambda_2 > \lambda_3 > 0$ for minima, $\lambda_1 > \lambda_2 > 0 > \lambda_3$ for pancake-like saddle points, $\lambda_1 > 0 > \lambda_2 > \lambda_3$ for filamentary saddle points and $0 > \lambda_1 > \lambda_2 > \lambda_3$ for maxima. The integral to perform is similar to that for the Euler characteristic, e.g. equation (31) in 3D,

$$n_{\mathrm{ext,3D}} = \frac{1}{\sigma_{1\perp}^2 \sigma_{1\parallel}} \int_{\mathrm{fixed}\lambda_i\mathrm{signs}} \mathrm{d}x\, \mathrm{d}\zeta\, \mathrm{d}J_{2\perp} \mathrm{d}\xi\, \mathrm{d}Q^2 \mathrm{d}\Upsilon\, P_{\mathrm{ext}}(x, \zeta, J_{2\perp}\xi, Q^2, \Upsilon)|I_3|, \tag{62}$$

the principal difference being in the limits of integration. The signature of the eigenvalues set changes where I_3 changes sign. In terms of the invariant variables, I_3 is given by equation (32). The equation $I_3 = 0$ that follows from equation (32) is more instructive if one writes it in the form that uses the full 3D Hessian rotation invariants, J_1, J_2, J_3,

$$J_1^3 - 3J_2\left(y, J_{2\perp}, Q^2, \Upsilon\right) J_1 - 2J_3\left(y, J_{2\perp}, Q^2, \Upsilon\right) = 0\,, \tag{63}$$

where $J_1 = J_{1\perp} + x_{33}, J_2 = y^2 + 3J_{2\perp}/4 + 3Q^2, J_3 = -y^3 + 9y(J_{2\perp} - 2Q^2)/4 + 27\Upsilon/4$ with $y = J_{1\perp}/2 - x_{33}$ where for this equation only, for the sake of simplicity, variables are not rescaled by their variance. It can be shown that for any values of $Q^2, J_{2\perp}, \Upsilon$ and y in their unrestricted allowed range, there exist three real roots for this cubic polynomial in J_1 that split the integration over J_1 in four regions corresponding to different extrema types. No restriction on other variables arises besides choosing the threshold of the field value for differential counts.

The integral required to predict extrema counts in 2D is given by equation (39) [where $P_{2\mathrm{D}}(\tilde{x}, \tilde{x}_1, \tilde{x}_2, \tilde{\zeta}, \tilde{\xi}, \tilde{x}_{12})$ is defined in equation (24)] and is carried out in terms of the field variables, $\tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{22}$, subject to constraints on the signs of the eigenvalues of the Hessian.

Extrema counts in anisotropic 2D and 3D spaces do not have closed-form expressions (indeed, differential 3D extrema counts do not lead to analytic results even in the Gaussian limit). Therefore, we shall not present here intermediate expressions for formal expansion, and instead will be performing the averaging numerically. Recall, however, that in the rare event limit, $\nu \gg 1$ or $\nu \ll -1$, $n_{\mathrm{max/min}}(\nu) \simeq \chi(\nu)$, so that in this limit, equations (40) and (C4) provide an all-order expansion of the extrema counts in redshift space in 2D and 3D, respectively.

We can also establish the following general symmetry relations between extrema counts of different types, valid in a space of arbitrary dimension N. Let us label the extrema type by its signature S which is the sum of the signs of the eigenvalues of the Hessian that define its type. In this way, $S = N$ for minima (N positive eigenvalues), $S = -N$ for maxima (N negative eigenvalues) and S changes with a step of two between $2 - N$ to $N - 2$ for saddle points of different types. Then, if we denote by $\partial_\nu n_{\mathrm{S}}^{(n)}$ the contribution to the differential number count of extrema of type S of order n (where $n = 2$ corresponds to the Gaussian term), the following holds

$$\partial_\nu n_{\mathrm{S}}^{(n)}(\nu) = (-1)^n \partial_\nu n_{-\mathrm{S}}^{(n)}(-\nu) \quad . \tag{64}$$

These relations allow us to predict the expected behaviour of extrema counts of type S from the measurements of their 'conjugate' type $-S$. In particular, the minima counts GC terms are equal to the reflected ($\nu \to -\nu$) maxima counts for even n (including the Gaussian term), and to minus the reflected maxima counts for odd n (including the first NG correction). The same relation holds between the 'pancake-like' ($S = 1$) and the 'filament-like' ($S = -1$) saddle points in 3D.

Fig. B4 illustrates the corresponding extrema distribution for a set of anisotropic fields (Gaussian and first-order NG correction) in 2D and 3D. A comparison with extrema counts measured from random realizations of the fields (scale-invariant $n = -1$ field sampled on 4096^2 pixels) is also presented there. Fig. 5 demonstrates the angular dependence of the predicted extrema counts at NLO in a scale-invariant LSS simulation ($n = -1$, $\sigma = 0.18$ in real space).

4.5 Invariance of critical sets

4.5.1 Summary of the Minkowski functionals as functions of ν_f

To put the mathematical results derived in Sections 4.2.1–4.3.3 on a practical footing in cosmology, let us collect them expressed as functions of the observable threshold variable ν_f. The $\nu \to \nu_f$ remapping is astrophysically motivated by the fact that typically the amplitude of the field is not known, as it may depend on e.g. the bias factor, whereas ν_f can be measured.[5] The value of ν_f is obtained by inverting equation (29) for the filling factor, f_V, and using it instead of the difficult-to-determine ν, to effectively 'Gaussianize' the PDF of the field. Indeed, when the transformation $\nu \to \nu_f$ in equation (30) is applied to perturbative results, the most oscillatory ν modes, which are proportional solely to the cumulants of the field, are eliminated.

In analogy with the skewness parameter $S_3 = \langle x^3\rangle/\sigma$, we introduce two other scaled cumulants that involve the derivatives of the field. For isotropic fields, they are $T_3 = \langle xq^2\rangle/\sigma$ and $U_3 = -\langle J_1 q^2\rangle/(\gamma\sigma)$. For anisotropic fields, we use their partial versions $T_{3\perp} = \langle xq_\perp^2\rangle/\sigma$, $T_{3\parallel} = \langle xq_\parallel^2\rangle/\sigma$ and $U_{3\perp} = -\langle J_{1\perp}q_\perp^2\rangle/(\gamma_\perp\sigma)$, $U_{3\parallel} = -\langle J_{1\perp}q_\parallel^2\rangle/(\gamma_\perp\sigma)$. In the isotropic limit, these partial contributions are summed according to the rules $T_3 = \frac{1}{3}(2T_{3\perp} + T_{3\parallel})$ and $U_3 = \frac{4}{9}(U_{3\perp} + U_{3\parallel})$.[6] It is then straightforward to rewrite our statistics, e.g. the 3D Euler characteristic, as a function of ν_f,

$$\chi_{3\mathrm{D}}(\nu_f) = \frac{\sigma_{1\parallel}\sigma_{1\perp}^2}{\sigma^3}\frac{\mathrm{e}^{-\nu_f^2/2}}{8\pi^2}\left(H_2(\nu_f) + \sigma\left[\left(T_{3\perp} + \frac{1}{2}T_{3\parallel} - S_3\right)H_3(\nu_f) + \left(U_{3\perp} + U_{3\parallel} - S_3\right)H_1(\nu_f)\right] + \cdots\right)\,. \tag{65}$$

Note that in contrast to the isotropic case, in redshift space the coefficient of H_1, $\tilde{S}_z^{(1)} = U_{3\perp} + U_{3\parallel} - S_3$, is non-zero for scale-invariant power spectra. For the other statistics, we get the following:

[5] Note that the choice of ν_f as an invariant parametrization of the Minkowski functionals is not unique. One could plot Minkowski as a function of e.g. ν_L, the threshold corresponding to a given fraction of the total skeleton length as an alternative construction.

[6] To see this, we use the relations between isotropic and anisotropic variances and correlation parameters from Appendix A1 as well as the following identities: $\gamma_\perp\langle J_{1\parallel}q_\perp^2\rangle = \gamma_\parallel\langle J_{1\perp}q_\parallel^2\rangle$ and $\langle J_{1\parallel}q_\parallel^2\rangle = 0$.

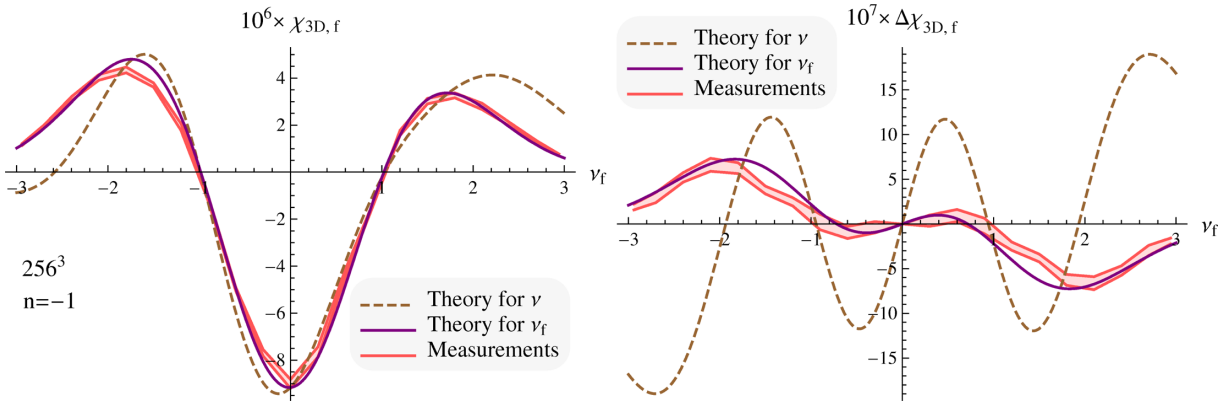


Figure 6. Left-hand panel : 3D Euler characteristic as a function of ν_f in 19 scale-invariant LSS simulations ($n=-1$, 256^3, $\sigma=0.18$). Prediction to first order in non-Gaussianity is displayed as a solid purple line. The displayed error is the error on the median of the measurements (red shaded area). For comparison, the 3D Euler characteristic as a function of ν is displayed as a dashed brown line. Right-hand panel: same as the left-hand panel for the first NG correction only compared to the odd part of the measurements (including the first-order correction). In this mildly non-linear regime, the agreement between measurements and theory at leading and NLO is very good. Note how the use of ν_f instead of ν 'Gaussianizes' the Euler characteristic.

the area of 3D isocontours

$$\mathcal{N}_3(\nu_f)=\mathcal{N}_3^{(0)}(\nu_f)\left(1+\sigma H_1(\nu_f)\left[\frac{1}{2}\left(1-\frac{A(\beta_\sigma)/\beta_\sigma}{1-A(\beta_\sigma)}\right)T_{3\perp}+\frac{1}{2}\left(\frac{A(\beta_\sigma)/\beta_\sigma}{1-A(\beta_\sigma)}\right)T_{3\parallel}-S_3\right]+\cdots\right), \tag{66}$$

the 2D Euler characteristic

$$\chi_{2D}(\nu_f,\theta_S)=\frac{\tilde\sigma_{1\perp}\tilde\sigma_{1\parallel}}{\sigma^2}\frac{\mathrm{e}^{-\nu_f^2/2}}{(2\pi)^{3/2}}\left(H_1\left(\nu_f\right)+\sigma\left[H_2(\nu_f)\left(\frac{1}{2}\frac{\left\langle\tilde x\left[\tilde x_1^2+\tilde x_2^2\right]\right\rangle}{\sigma}-S_3\right)+\left(-\frac{1}{2\tilde\gamma_\perp}\frac{\left\langle\tilde x_{11}\tilde x_2^2\right\rangle}{\sigma}-S_3\right)\right]+\cdots\right), \tag{67}$$

the length of 2D isocontours

$$\mathcal{N}_2(\nu_f,\theta_S)=\mathcal{N}_2^{(0)}(\nu_f)\left(1+\sigma H_1(\nu_f)\left[\frac{1}{2}\frac{\left\langle\tilde x\tilde x_1^2\right\rangle}{\sigma}B(\tilde\beta_\sigma)+\frac{1}{2}\frac{\left\langle\tilde x\tilde x_2^2\right\rangle}{\sigma}\left(1-B(\tilde\beta_\sigma)\right)-S_3\right]+\cdots\right) \tag{68}$$

and the frequency of contour crossings

$$\mathcal{N}_1(\nu_f,\theta_\mathcal{L})=\mathcal{N}_1^{(0)}(\nu_f)\left(1+\sigma H_1(\nu_f)\left[\frac{1}{2}\frac{(1-\beta_\sigma)\sin^2\theta_\mathcal{L}T_{3\perp}+\cos^2\theta_\mathcal{L}T_{3\parallel}}{1-\beta_\sigma\sin^2\theta_\mathcal{L}}-S_3\right]+\cdots\right). \tag{69}$$

Note that $\langle\tilde x\tilde x_1^2\rangle$, $\langle\tilde x\tilde x_1^2\rangle$ and $\langle\tilde x_{11}\tilde x_2^2\rangle$ are functions of $\theta_{S/\mathcal{L}}$, $T_{3\perp}$, $T_{3\parallel}$, $U_{3\perp}$ and $U_{3\parallel}$. Using ν_f as a threshold variable makes explicit the invariance of Minkowski functionals and extrema statistics under a monotonic transformation of the field. Indeed, since the filling factor f_V is one of the Minkowski functionals itself, ν_f is strictly unchanged under the monotonic transformation. Thus, the other statistics, described by functions of ν_f, are invariant.

Fig. 6 reproduces the 3D Euler characteristic of Fig. 3 but as a function of the filling factor threshold, ν_f, defined in Section 4.1. Agreement between the prediction truncated at NLO and measurements is very good for contrasts $\nu\gtrsim-1.5$. The effect of using ν_f is to Gaussianize the PDF as seen in Fig. 6: denser regions are brought to the centre whereas wider regions are pushed to the outside.

4.5.2 Formal invariance of Minkowski functionals w.r.t. monotonic transformation

From Section 2, it is straightforward to see that any Minkowski functional as well as extrema counts are invariant under a local monotonic transformation. This property should also be encoded in the GC expansions of these Minkowski functionals given in Section 4.5.1. Indeed, the combinations of cumulants which appear at first order in the GC expansion are invariant under a local monotonic transformation taken to the same first order in σ. Let us illustrate this on the 3D Euler characteristic (given that the same proof can be developed for any Minkowski functionals). Following Matsubara (2003), let us henceforth study how cumulants evolve under the transformation: $y\to b\,x+b_2/2(x^2-\langle x^2\rangle)$, which represents the local representation of any analytic function of the density field when Taylor expanded. Let us first show how $\langle x^3\rangle$ evolve under such a transformation,

$$\begin{aligned}\left\langle y^3\right\rangle&=b^3\left\langle x^3\right\rangle+\frac{3}{2}b^2b_2\left(\left\langle x^4\right\rangle-\left\langle x^2\right\rangle^2\right)+\frac{3}{4}b\,b_2^2\left(\left\langle x^5\right\rangle-2\left\langle x^3\right\rangle\left\langle x^2\right\rangle\right)\\&=b^3\left\langle x^3\right\rangle_c+\frac{3}{2}b^2b_2\left(\left\langle x^4\right\rangle_c+2\left\langle x^2\right\rangle_c^2\right)+\frac{3}{4}b\,b_2^2\left(\left\langle x^5\right\rangle_c+8\left\langle x^3\right\rangle_c\left\langle x^2\right\rangle_c\right).\end{aligned} \tag{70}$$

Selecting only the first-order term in the PT expansion, the classical expression for the skewness follows from equation (70),

$$\left\langle y^3\right\rangle=b^3\left\langle x^3\right\rangle_c+3b^2b_2\left\langle x^2\right\rangle_c^2+\text{higher order terms}. \tag{71}$$

The same construction for other moments and cumulants leads to the following relationships:

$$\left\langle y^2\right\rangle = b^2\left\langle x^2\right\rangle + \text{higher order terms}, \tag{72a}$$

$$\langle \nabla_\perp y\cdot\nabla_\perp y\rangle = b^2\langle \nabla_\perp x\cdot\nabla_\perp x\rangle + \text{higher order terms}, \tag{72b}$$

$$\left\langle \nabla_\parallel y\cdot\nabla_\parallel y\right\rangle = b^2\left\langle \nabla_\parallel x\cdot\nabla_\parallel x\right\rangle + \text{higher order terms}, \tag{72c}$$

$$\left\langle y(\nabla_\perp y)^2\right\rangle = b^3\left\langle x(\nabla_\perp x)^2\right\rangle + 2b^2b_2\left\langle x^2\right\rangle\left\langle(\nabla_\perp x)^2\right\rangle + \text{higher order terms}, \tag{72d}$$

$$\left\langle y(\nabla_\parallel y)^2\right\rangle = b^3\left\langle x(\nabla_\parallel x)^2\right\rangle + 2b^2b_2\left\langle x^2\right\rangle\left\langle(\nabla_\parallel x)^2\right\rangle + \text{higher order terms}, \tag{72e}$$

$$\langle(\nabla_\perp y\cdot\nabla_\perp y)\Delta_\perp y\rangle = b^3\langle(\nabla_\perp x\cdot\nabla_\perp x)\Delta_\perp y\rangle - b^2b_2\left\langle(\nabla_\perp x)^2\right\rangle^2 + \text{higher order terms}, \tag{72f}$$

$$\left\langle(\nabla_\parallel y\cdot\nabla_\parallel y)\Delta_\perp y\right\rangle = b^3\left\langle(\nabla_\parallel x\cdot\nabla_\parallel x)\Delta_\perp y\right\rangle - 2b^2b_2\left\langle(\nabla_\perp x)^2\right\rangle\left\langle(\nabla_\parallel x)^2\right\rangle + \text{higher order terms}, \tag{72g}$$

so that for $b\neq 0$ at leading order,

$$\tilde{S}^{(1)}_{\chi_{3D},z}(y) \equiv \frac{1}{\left\langle y^2\right\rangle}\left(\frac{\left\langle y(\nabla_\perp y)^2\right\rangle}{\left\langle(\nabla_\perp y)^2\right\rangle} + \frac{\left\langle y(\nabla_\parallel y)^2\right\rangle}{2\left\langle(\nabla_\parallel y)^2\right\rangle} - \frac{\left\langle y^3\right\rangle}{\left\langle y^2\right\rangle}\right) = \tilde{S}^{(1)}_{\chi_{3D},z}(x)\,, \tag{73}$$

$$\tilde{S}^{(-1)}_{\chi_{3D},z}(y) \equiv \frac{1}{\left\langle y^2\right\rangle}\left(-\frac{\left\langle \Delta_\perp y(\nabla_\perp y)^2\right\rangle}{\left\langle(\nabla_\perp y)^2\right\rangle} - \frac{\left\langle \Delta_\perp y(\nabla_\parallel y)^2\right\rangle}{\left\langle(\nabla_\parallel y)^2\right\rangle} - \frac{\left\langle y^3\right\rangle}{\left\langle y^2\right\rangle}\right) = \tilde{S}^{(-1)}_{\chi_{3D},z}(x)\,, \tag{74}$$

where we denote by $\tilde{S}^{(i)}_{\chi_{3D},z}$ the coefficients in front of the Hermite polynomials in equation (65) [see also equation (25)]. This shows that the combinations of cumulants in front of the Hermite polynomials are invariant under a local monotonic transformation and so are the Minkowski functionals. In particular, it demonstrates formally why if the bias is local and monotonic, the Minkowski functionals are bias independent.

5 APPLICATION TO LSS IN REDSHIFT SPACE

In the context of cosmology, it is of interest to understand what kind of constraints on the cosmological parameters can be drawn from the prediction of critical sets in redshift space. Redshift-space distortion can be viewed as a nuisance, but in fact potentially opens new prospects given the broken symmetry induced by the kinematics.

5.1 Estimating cumulants of the field

Table 1 summarizes the cubic cumulants that determine the Minkowski functionals studied in this paper to first NG order. In turn, the combinations of these cumulants are what can be measured by fitting the correspondent functionals with low-order Hermite modes. Two main groups of cumulants that Minkowski functionals (to first order) give access to are $\langle xq_\perp^2\rangle$, $\langle xx_3^2\rangle$, which relate the field to its gradient, and $\langle J_{1\perp}q_\perp^2\rangle$, $\langle J_{1\perp}x_3^2\rangle$ which relate the gradient to the Hessian. Note that $\langle x^3\rangle$ is not accessible as the amplitude of an independent Hermite mode if Minkowski functionals are studied as functions of the filling factor threshold ν_f. Rather it offsets the modes defined by the other two groups of cumulants. Redshift-space analysis is in principle capable of mining more information than real-space analysis. Indeed, in redshift space there is a qualitative difference between cumulants that involve the LOS direction and those that involve directions orthogonal to the LOS in the plane of the sky. These differences encode information about velocities, and reflect the mechanism of how these velocities originated. In principle, estimating the anisotropic part of such cumulants can be used to test the theory of gravity in the context of LSS PT. 3D geometrical statistics, such as χ_{3D} and $\mathcal{N}_3$, do not allow by themselves to determine separately the LOS and sky cumulants. To separate

Table 1. The cubic cumulants that determine the Minkowski functionals studied in this paper to first NG order.

	H_0	H_1	H_2	H_3
$\chi_{3D,\,f}$	–	$\langle J_{1\perp}q_\perp^2\rangle$, $\langle J_{1\perp}x_3^2\rangle$, $\langle x^3\rangle$	–	$\langle xq_\perp^2\rangle$, $\langle xx_3^2\rangle$, $\langle x^3\rangle$
$\chi_{2D,\,f}$	$\langle J_{1\perp}q_\perp^2\rangle$, $\langle J_{1\perp}x_3^2\rangle$, $\langle x^3\rangle$	–	$\langle xq_\perp^2\rangle$, $\langle xx_3^2\rangle$, $\langle x^3\rangle$	–
$\mathcal{N}_{3,f}$	–	$\langle xq_\perp^2\rangle$, $\langle xx_3^2\rangle$, $\langle x^3\rangle$	–	–
$\mathcal{N}_{2,f}$	–	$\langle xq_\perp^2\rangle$, $\langle xx_3^2\rangle$, $\langle x^3\rangle$	–	–
$\mathcal{N}_{1,f}$	–	$\langle xq_\perp^2\rangle$, $\langle xx_3^2\rangle$, $\langle x^3\rangle$	–	–

Table 2. Rescaled third-order cumulants written in terms of the bispectrum.

$\langle x^3\rangle$	$\frac{1}{\sigma^3}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}B(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)$
$\langle xq_\perp^2\rangle$	$\frac{1}{2\sigma\sigma_{1\perp}^2}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\lvert\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp}\rvert^2B(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)$
$\langle xq_\parallel^2\rangle$	$\frac{1}{2\sigma\sigma_{1\parallel}^2}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\lvert\boldsymbol{k}_{1\parallel}+\boldsymbol{k}_{2\parallel}\rvert^2B(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)$
$\langle J_{1\perp}q_\perp^2\rangle$	$-\frac{1}{\sigma_{2\perp}\sigma_{1\perp}^2}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\lvert\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp}\rvert^2(\boldsymbol{k}_{1\perp}\cdot\boldsymbol{k}_{2\perp})B(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)$
$\langle J_{1\perp}q_\parallel^2\rangle$	$-\frac{1}{\sigma_{2\perp}\sigma_{1\parallel}^2}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\lvert\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp}\rvert^2(\boldsymbol{k}_{1\parallel}\cdot\boldsymbol{k}_{2\parallel})B(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)$

anisotropic contributions, one must analyse *slices* of 3D volume at different angles θ_S to the LOS.[7] For instance, measuring the length of the isocontours, $\mathcal{N}_2$ (with possible cross-check from $\chi_{2\mathrm{D}}$), yields a separate handle on $\langle xq_\perp^2\rangle$ and $\langle xx_3^2\rangle$, while the additional analysis of the 2D Euler characteristic $\chi_{2\mathrm{D}}(\nu_f)$, as a function of θ_S (via the baseline offset $\propto H_0$, see Fig. 5), allows us to measure $\langle J_{1\perp}q_\perp^2\rangle$ and $\langle J_{1\perp}x_3^2\rangle$. This procedure is further discussed in Section 5.5 below.

5.2 Which modes of the bispectrum are geometrical statistics probing?

This paper is concerned with geometric probes operating in configuration space. On the other hand, a fair amount of theoretical predictions (such as PT) for the growth of structure are best described in Fourier space. It is therefore of interest to relate the two and characterize which feature of the multispectra these probes constrain. For instance, it is straightforward to show via Fourier transform that the third-order cumulant, $\langle x^3\rangle$, can be expressed as a double sum over the anisotropic bispectrum, $B_z(k_1,k_2,k_3)$, via

$$\left\langle x^3\right\rangle=\frac{1}{\sigma^3}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}B(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)\,.$$

Measuring $\langle x^3\rangle$ amounts to constraining the monopole of the bispectrum. Similarly, other geometric cumulants involve different weights (see Table 2), while higher order cumulants will involve k integrals of multispectra. For instance, the Euler characteristic to all orders given in equation (C5) involves $n\mathrm{th}=i+2j+2m$ moments, $\langle x^iq_\perp^{2j}x_3^{2m}\rangle$, which can be re-expressed via the nth-order multispectrum, $B_z^n(k_1,\ldots,k_{n-1},\lvert\boldsymbol{k}_1+\cdots\boldsymbol{k}_{n-1}\rvert)$, as

$$\left\langle x^iq_\perp^{2j}x_3^{2m}\right\rangle=\frac{1}{\sigma^i\sigma_{1\perp}^{2j}\sigma_{1\parallel}^{2m}}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\cdots\frac{\mathrm{d}^3k_{n-1}}{(2\pi)^3}B_z^n(k_1,\ldots,k_{n-1},\lvert\boldsymbol{k}_1+\cdots\boldsymbol{k}_{n-1}\rvert)\prod_{j_1\leq j}\lvert\boldsymbol{k}_{j_1\perp}\rvert^2\prod_{j<m_1\leq m+j}\lvert\boldsymbol{k}_{m_1\parallel}\rvert^2\,.$$

Let us first show explicitly how the first-order corrections of the 3D Euler characteristic can be re-expressed in terms of the underlying bispectrum in redshift space, B_z. Equation (65) shows that it only depends at first order in non-Gaussianity on two numbers (the coefficients in front of the two Hermite polynomials): $\sigma\tilde{S}^{(1)}_{\chi_{3\mathrm{D}},z}=\langle x(q_\perp^2+x_3^2/2)\rangle-\langle x^3\rangle$ and $\sigma\tilde{S}^{(-1)}_{\chi_{3\mathrm{D}},z}=-\langle J_{1\perp}(q_\perp^2+x_3^2)\rangle/\gamma_\perp-\langle x^3\rangle$. These quantities can in turn be expressed as special combinations of the underlying bispectrum using Table 2,

$$\tilde{S}^{(1)}_{\chi_{3\mathrm{D}},z}=\frac{1}{\sigma^4}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\left(\frac{\sigma^2\lvert\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp}\rvert^2}{2\sigma_{1\perp}^2}+\frac{\sigma^2\lvert\boldsymbol{k}_{1\parallel}+\boldsymbol{k}_{2\parallel}\rvert^2}{2\sigma_{1\parallel}^2}-1\right)B_z(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)\,,\tag{75}$$

$$\tilde{S}^{(-1)}_{\chi_{3\mathrm{D}},z}=-\frac{1}{\sigma^4}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\left(\frac{\sigma^4\lvert\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp}\rvert^2(\boldsymbol{k}_{1\perp}\cdot\boldsymbol{k}_{2\perp})}{\sigma_{1\perp}^4}+\frac{\sigma^4\lvert\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp}\rvert^2(\boldsymbol{k}_{1\parallel}\cdot\boldsymbol{k}_{2\parallel})}{\sigma_{1\perp}^2\sigma_{1\parallel}^2}+1\right)B_z(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)\,.\tag{76}$$

The parentheses in equations (75) and (76) define 'projectors' for the bispectrum, $(\tilde{S}^{(1)}_{\chi_{3\mathrm{D}},z})_B$ and $(\tilde{S}^{(-1)}_{\chi_{3\mathrm{D}},z})_B$, so that

$$\tilde{S}^{(i)}_{\chi_{3\mathrm{D}},z}=\frac{1}{\sigma^4}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\,B_z(k_1,k_2,\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert)\left(\tilde{S}^{(i)}_{\chi_{3\mathrm{D}},z}\right)_B\,.$$

In the isotropic limit, these projectors become, respectively (Matsubara 2003),

$$\left(\tilde{S}^{(1)}_{\chi_{3\mathrm{D}},z}\right)_B=\left(\frac{3}{2}\frac{\sigma^2\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert^2}{2\sigma_1^2}-1\right)\quad\text{and}\quad\left(\tilde{S}^{(-1)}_{\chi_{3\mathrm{D}},z}\right)_B=\left(\frac{9}{4}\frac{\sigma^4\lvert\boldsymbol{k}_1+\boldsymbol{k}_2\rvert^2(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)}{\sigma_1^4}+1\right)\,.\tag{77}$$

[7] Another related approach would be to study the 2D slices orthogonal to LOS field with variable LOS thickness. This latter technique was successfully used to study ISM turbulence (Lazarian & Pogosyan 2000).

A multipole expansion of B_z and equation (77) w.r.t. $\mu = \boldsymbol{k}_1 \cdot \boldsymbol{k}_2/(k_1 k_2)$ show for instance that neither $\tilde{S}^{(1)}_{\chi 3\mathrm{D}}$ nor $\tilde{S}^{(-1)}_{\chi 3\mathrm{D}}$ would constrain harmonics of B_z larger than three. For the 2D Euler characteristic, the corresponding projectors read [with β_σ defined in equation (26)]

$$\left(\tilde{S}^{(1)}_{\chi 2\mathrm{D},z}\right)_B = \left(1 - \frac{\cos^2\theta_{\mathcal{S}}}{2(1-\beta_\sigma \sin^2\theta_{\mathcal{S}})}\right) \frac{\sigma^2 |\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2}{2\sigma^2_{1\perp}} + \frac{\cos^2\theta_{\mathcal{S}}}{2(1-\beta_\sigma \sin^2\theta_{\mathcal{S}})} \frac{\sigma^2 |\boldsymbol{k}_{1\parallel} + \boldsymbol{k}_{2\parallel}|^2}{2\sigma^2_{1\parallel}} - 1\,, \tag{78}$$

$$\left(\tilde{S}^{(-1)}_{\chi 2\mathrm{D},z}\right)_B = \left(1 - \frac{\cos^2\theta_{\mathcal{S}}}{1-\beta_\sigma \sin^2\theta_{\mathcal{S}}}\right) \frac{\sigma^4 |\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 (\boldsymbol{k}_{1\perp} \cdot \boldsymbol{k}_{2\perp})}{\sigma^4_{1\perp}} + \frac{\cos^2\theta_{\mathcal{S}}}{2(1-\beta_\sigma \sin^2\theta_{\mathcal{S}})} \frac{\sigma^4 |\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 (\boldsymbol{k}_{1\parallel} \cdot \boldsymbol{k}_{2\parallel})}{\sigma^2_{1\perp}\sigma^2_{1\parallel}} - 1\,. \tag{79}$$

An interesting feature of equations (75) and (76) is that the projectors are now parametric and depend on $\theta_{\mathcal{S}}$. Via slicing planes e.g. along and perpendicular to the LOS, it is therefore possible to measure projections of the bispectrum along $|\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2$, $|\boldsymbol{k}_{1\parallel} + \boldsymbol{k}_{2\parallel}|^2$, $|\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 (\boldsymbol{k}_{1\perp} \cdot \boldsymbol{k}_{2\perp})$ and $|\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 (\boldsymbol{k}_{1\parallel} \cdot \boldsymbol{k}_{2\parallel})$ independently.

Given equations (66), (68) and (69), one can also easily recover the projectors for $\mathcal{N}_{1,f}$, $\mathcal{N}_{2,f}$ and $\mathcal{N}_{3,f}$ at first order in non-Gaussianity as

$$\left(\tilde{S}^{(1)}_{N_3,z}\right)_B = \frac{\sigma^2}{4\sigma^2_{1\perp}} \left(1 - \frac{A(\beta_\sigma)/\beta_\sigma}{1 - A(\beta_\sigma)}\right) |\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 + \frac{\sigma^2}{4\sigma^2_{1\parallel}} \left(\frac{A(\beta_\sigma)/\beta_\sigma}{1 - A(\beta_\sigma)}\right) |\boldsymbol{k}_{1\parallel} + \boldsymbol{k}_{2\parallel}|^2 - 1\,, \tag{80}$$

$$\left(\tilde{S}^{(1)}_{N_2,z}(\theta_{\mathcal{S}})\right)_B = \frac{\sigma^2}{4\sigma^2_{1\perp}} B(\tilde{\beta}_\sigma) |\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 + \frac{1}{4}\sigma^2 \frac{\sin^2\theta_{\mathcal{S}} |\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 + 2\cos^2\theta_{\mathcal{S}} |\boldsymbol{k}_{1\parallel} + \boldsymbol{k}_{2\parallel}|^2}{\sin^2\theta_{\mathcal{S}} \sigma^2_{1\perp} + 2\cos^2\theta_{\mathcal{S}} \sigma^2_{1\parallel}} \left(1 - B(\tilde{\beta}_\sigma)\right) - 1\,, \tag{81}$$

$$\left(\tilde{S}^{(1)}_{N_1,z}(\theta_{\mathcal{L}})\right)_B = \frac{1}{4} \frac{(1-\beta_\sigma)\sin^2\theta_{\mathcal{L}} |\boldsymbol{k}_{1\perp} + \boldsymbol{k}_{2\perp}|^2 \sigma^2/\sigma^2_{1\perp} + \cos^2\theta_{\mathcal{L}} |\boldsymbol{k}_{1\parallel} + \boldsymbol{k}_{2\parallel}|^2 \sigma^2/\sigma^2_{1\parallel}}{1 - \beta_\sigma \sin^2\theta_{\mathcal{L}}} - 1\,, \tag{82}$$

where the last two are also parametric in $\theta_{\mathcal{L}}$ and $\theta_{\mathcal{S}}$, respectively. Note that equations (80)–(82) formally yield no new projector, compared to equation (78), though the weighting differs and might be more favourable for noisy data sets. As no closed form for the extrema counts was found, it is not possible to extend this analysis to their cumulants.

5.3 Predicting cumulants using gravitational PT

In the previous section, no assumption was made on the shape of the anisotropic bispectrum, B_z. Let us now turn to the context of gravitational clustering in redshift space and start with a rapid overview of the relevant theory. The fully non-linear expression [generalizing equation (11)] for the Fourier transform of the density in redshift space is (Scoccimarro et al. 1999; Bernardeau et al. 2002)

$$\hat{\delta}_s(\boldsymbol{k}) = \int \frac{\mathrm{d}^3\boldsymbol{x}}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{e}^{\mathrm{i} f k_z v_z(\boldsymbol{x})} \left[\delta^{(r)}(\boldsymbol{x}) + f\nabla_z v_z(\boldsymbol{x})\right]\,, \tag{83}$$

with v_z the peculiar velocity along the LOS, $f = \mathrm{d}\log D/\mathrm{d}\log a$, while assuming the plane-parallel approximation and that only $f\nabla_z v_z(\boldsymbol{x}) < 1$ terms contribute.[8]

5.3.1 Derivation of geometrical cumulants for standard gravitational clustering

Expanding the exponential in equation (83) leads to, using the kernels Z_n, the following expression for the density field in redshift space (Verde et al. 1998; Scoccimarro et al. 1999; Bernardeau et al. 2002)

$$\hat{\delta}_s(\boldsymbol{k},\tau) = \sum_{n=1}^{\infty} D_1^n(\tau) \int \mathrm{d}^3\boldsymbol{k}_1 \cdots \int \mathrm{d}^3\boldsymbol{k}_n \delta_{\mathrm{D}}(\boldsymbol{k} - \boldsymbol{k}_1 - \cdots - \boldsymbol{k}_n) Z_n(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_n) \hat{\delta}_l(\boldsymbol{k}_1) \cdots \hat{\delta}_l(\boldsymbol{k}_n), \tag{84}$$

where μ_i is the cosine of the angle between $\boldsymbol{k}_i$ and the LOS, $\boldsymbol{k} = \boldsymbol{k}_1 + \boldsymbol{k}_2$. The first kernels are given by (assuming a quadratic local bias model involving b_1 and b_2)

$$Z_1(\boldsymbol{k}) = (b_1 + f\mu^2) \;\text{ and }\; Z_2(\boldsymbol{k}_1,\boldsymbol{k}_2) = b_1 F_2(\boldsymbol{k}_1,\boldsymbol{k}_2) + f\mu^2 G_2(\boldsymbol{k}_1,\boldsymbol{k}_2) + \frac{f\mu k}{2}\left[\frac{\mu_1}{k_1}\left(b_1 + f\mu_2^2\right) + \frac{\mu_2}{k_2}\left(b_1 + f\mu_1^2\right)\right] + \frac{b_2}{2}\,, \tag{85}$$

with $f = \Omega_{\mathrm{m}}^{\gamma_m}$, $\gamma_{\mathrm{m}} \simeq 6/11$, $\epsilon \simeq 3/7\,\Omega_{\mathrm{m}}^{-2/63}$ and

$$F_2(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{1}{2}(1+\epsilon) + \frac{1}{2}\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{1}{2}(1-\epsilon)\left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1 k_2}\right)^2\,, \tag{86}$$

[8] Note that equation (83) could be more accurately replaced by $\hat{\delta}_s(\boldsymbol{k}) = -\delta_{\mathrm{D}}(\boldsymbol{k}) + 1/(2\pi)^3 \int \mathrm{d}^3\boldsymbol{x} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{e}^{\mathrm{i} f k_z v_z(\boldsymbol{x})} [\delta^{(r)}(\boldsymbol{x}) + 1]$, where no assumption about the amplitude of the radial velocity is made. This gives exactly the same PT as expected.

Table 3. α and β coefficients defined via equation (90) for the relevant cumulants.

	α	β
$\langle\delta^3\rangle$	3	3
$\langle\delta^2 J_{1\perp}\rangle$	$-2k_{1\perp}^2$	$-(\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp})^2-2(k_{2\perp})^2$
$\langle\delta q_\perp^2\rangle$	$k_{1\perp}^2$	$\boldsymbol{k}_{1\perp}\cdot\boldsymbol{k}_{2\perp}+2k_{1\perp}^2$
$\langle\delta q_\parallel^2\rangle$	$k_{1\parallel}^2$	$k_{1\parallel}k_{2\parallel}+2k_{1\parallel}^2$
$\langle\delta I_{2\perp}\rangle$	0	$\frac{3}{4}\left((\boldsymbol{k}_{1\perp}\times\boldsymbol{k}_{2\perp})^2-(\boldsymbol{k}_{1\perp}\cdot\boldsymbol{k}_{2\perp})^2+k_{1\perp}^2k_{2\perp}^2\right)$
$\langle J_{1\perp}q_\perp^2\rangle$	0	$-(\boldsymbol{k}_{1\perp}\times\boldsymbol{k}_{2\perp})^2+(\boldsymbol{k}_{1\perp}\cdot\boldsymbol{k}_{2\perp})^2-k_{1\perp}^2k_{2\perp}^2$
$\langle J_{1\perp}q_\parallel^2\rangle$	0	$k_{1z}k_{2z}(\boldsymbol{k}_{1\perp}+\boldsymbol{k}_{2\perp})^2-2k_{1\perp}^2k_{2z}(k_{1z}+k_{2z})$

$$G_2(\boldsymbol{k}_1,\boldsymbol{k}_2)=\epsilon+\frac{1}{2}\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)+(1-\epsilon)\left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}\right)^2. \tag{87}$$

In particular

$$\hat{\delta}_s^{(1)}(\boldsymbol{k})=D_1(\tau)\int \mathrm{d}^3\boldsymbol{k}_1\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}_1)Z_1(\boldsymbol{k}_1)\hat{\delta}_l(\boldsymbol{k}_1)=D_1(\tau)Z_1(\boldsymbol{k})\hat{\delta}_l(\boldsymbol{k}), \tag{88}$$

$$\hat{\delta}_s^{(2)}(\boldsymbol{k})=D_1^2(\tau)\int \mathrm{d}^3\boldsymbol{k}_1\int \mathrm{d}^3\boldsymbol{k}_2\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}_1-\boldsymbol{k}_2)Z_2(\boldsymbol{k}_1,\boldsymbol{k}_2)\hat{\delta}_l(\boldsymbol{k}_1)\hat{\delta}_l(\boldsymbol{k}_2). \tag{89}$$

Given this expansion, cumulants can be computed at tree order. For conciseness, we denote by simply $\int_W$ the weighted 6D integration $\int \mathrm{d}^3\boldsymbol{k}_i\mathrm{d}^3\boldsymbol{k}_j[\quad\;]W(k_iR)W(k_jR)W(|\boldsymbol{k}_i+\boldsymbol{k}_j|R)$. With this notation, the cumulant of δ^3 reads for instance

$$\begin{aligned}\left\langle\delta^3\right\rangle &\simeq 3\left\langle(\delta^{(1)})^2\delta^{(2)}\right\rangle = 3D_1^4(\tau)\int_W Z_1(\boldsymbol{k}_1)Z_1(\boldsymbol{k}_2)Z_2(\boldsymbol{k}_3,\boldsymbol{k}_4)\left\langle\hat{\delta}_l(\boldsymbol{k}_1)\hat{\delta}_l(\boldsymbol{k}_2)\hat{\delta}_l(\boldsymbol{k}_3)\hat{\delta}_l(\boldsymbol{k}_4)\right\rangle,\\ &= 3D_1^4(\tau)\int_W Z_1(\boldsymbol{k}_1)Z_1(\boldsymbol{k}_2)Z_2(\boldsymbol{k}_3,\boldsymbol{k}_4)\left(\left\langle\hat{\delta}_l(\boldsymbol{k}_1)\hat{\delta}_l(\boldsymbol{k}_2)\right\rangle\left\langle\hat{\delta}_l(\boldsymbol{k}_3)\hat{\delta}_l(\boldsymbol{k}_4)\right\rangle+2\left\langle\hat{\delta}_l(\boldsymbol{k}_1)\hat{\delta}_l(\boldsymbol{k}_3)\right\rangle\left\langle\hat{\delta}_l(\boldsymbol{k}_2)\hat{\delta}_l(\boldsymbol{k}_4)\right\rangle\right),\\ &= 3(2\pi)^6D_1^4(\tau)\left[\int_W Z_1(\boldsymbol{k}_1)Z_1(-\boldsymbol{k}_1)Z_2(\boldsymbol{k}_3,-\boldsymbol{k}_3)P(k_1)P(k_3)+2\int_W Z_1(\boldsymbol{k}_1)Z_1(\boldsymbol{k}_2)Z_2(-\boldsymbol{k}_1,-\boldsymbol{k}_2)P(k_1)P(k_2)\right],\end{aligned}$$

where we can note that $Z_2(\boldsymbol{k}_3,-\boldsymbol{k}_3)=b_2/2$. This method can be generalized to all relevant cumulants. Let us sum up these results in Table 3,[9] where each cumulant is generically written as

$$\langle\mathcal{Y}\rangle=(2\pi)^6D_1^4(\tau)\left[\int_W Z_1(\boldsymbol{k}_1)^2Z_2(\boldsymbol{k}_2,-\boldsymbol{k}_2)P(k_1)P(k_2)\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2)+2\int_W Z_1(\boldsymbol{k}_1)Z_1(\boldsymbol{k}_2)Z_2(\boldsymbol{k}_1,\boldsymbol{k}_2)P(k_1)P(k_2)\beta(\boldsymbol{k}_1,\boldsymbol{k}_2)\right], \tag{90}$$

with $\mathcal{Y}=\delta^3,\delta^2J_{1\perp},\delta q_\perp^2,\delta q_\parallel^2,\delta I_{2\perp},J_{1\perp}q_\perp^2$ and $J_{1\perp}q_\parallel^2$, respectively. As discussed in the previous section, the Minkowski functionals and the critical sets described in the main text yield access to (geometrically weighted by α and β) averages of products of the Z_1 and Z_2 kernels, which in turn depend on the underlying cosmological parameters via, say, ϵ in equations (85)–(87). Hence, provided the corresponding components of the Z_i's do not fall into the null space of the α and β projectors, one should expect to be able to access the values of some of these cosmic parameters through appropriate combinations of the geometrical sets, $\langle\mathcal{Y}\rangle=\langle\mathcal{Y}\rangle[\Omega_{\mathrm{m}},b_1,b_2,\gamma_{\mathrm{m}},D(z)]$. In practice, all these moments can be integrated using the decomposition of $W_G(|\boldsymbol{k}_1+\boldsymbol{k}_2|R)$ in Legendre polynomials and Bessel functions,

$$W_G(|\boldsymbol{k}_1+\boldsymbol{k}_2|R)=\exp\left[-\frac{k_1^2+k_2^2}{2}R^2\right]\sum_{l=0}^{\infty}(-1)^l(2l+1)P_l\left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}\right)I_{l+1/2}\left(k_1k_2R^2\right)\sqrt{\frac{\pi}{2k_1k_2R^2}}, \tag{91}$$

except one, given by

$$I=\int_W Z_1(\boldsymbol{k}_1)Z_1(\boldsymbol{k}_2)f\mu^2G_2(\boldsymbol{k}_1,\boldsymbol{k}_2)P(k_1)P(k_2)\beta(\boldsymbol{k}_1,\boldsymbol{k}_2). \tag{92}$$

This difficulty was highlighted e.g. by Hivon et al. (1995) but was not analytically solved until now. For this purpose, let us use the following trick: if we introduce three different smoothing lengths $W(k_1R_1)$, $W(k_2R_2)$ and $W(kR)$ with a Gaussian filter, then one can see that

$$-2\frac{\partial I}{\partial R^2}=\int_W Z_1(\boldsymbol{k}_1)Z_1(\boldsymbol{k}_2)f\mu^2k^2G_2(\boldsymbol{k}_1,\boldsymbol{k}_2)P(k_1)P(k_2)\beta(\boldsymbol{k}_1,\boldsymbol{k}_2). \tag{93}$$

[9] Note that these expressions (precisely, the second term in β) can be compared to the results of Gay et al. (2012) in real space. They are found to be in full agreement. In this table, it is also of interest to notice that $\langle\delta I_{2\perp}\rangle=-3/4\langle J_{1\perp}q_\perp^2\rangle$ as mentioned in Appendix A2.

Table 4. Predicted σ's and three-point cumulants for power-law power spectra in redshift space using PT.

	n = 1	n = 0	n = −1	n = −2
$\langle x^3\rangle/\sigma$	2.65	2.91	3.36	4.03
$\langle x^2 J_{1\perp}\rangle/\sigma$	−2.62	−2.59	−2.62	−2.47
$\langle x q_\perp^2\rangle/\sigma$	1.87	1.95	2.15	2.49
$\langle x q_\parallel^2\rangle/\sigma$	1.66	1.94	2.36	2.93
$\langle x I_{2\perp}\rangle/\sigma$	0.49	0.41	0.33	0.22
$\langle J_{1\perp} q_\perp^2\rangle/\sigma$	−0.94	−0.81	−0.72	−0.59
$\langle J_{1\perp} q_\parallel^2\rangle/\sigma$	−1.73	−1.54	−1.38	−1.16

Table 5. Predicted σ's and three-point cumulants for ΛCDM power spectra in redshift space using PT.

R (Mpc h^{-1}):	8	16	24	32	40	48	56	64
σ	0.54	0.29	0.19	0.13	0.10	0.081	0.066	0.055
$\sigma_{1\perp}$	0.047	0.014	0.0063	0.0035	0.0022	0.0015	0.0011	0.000 79
$\sigma_{1\parallel}$	0.038	0.011	0.0052	0.0029	0.0018	0.0012	0.000 88	0.000 65
$\sigma_{2\perp}$	0.0065	0.0010	0.000 32	0.000 14	0.000 070	0.000 040	0.000 025	0.000 016
$\sigma_{2\parallel}$	0.0050	0.000 78	0.000 25	0.000 11	0.000 053	0.000 030	0.000 019	0.000 012
$\langle x^3\rangle/\sigma$	3.68	3.49	3.39	3.34	3.30	3.26	3.22	3.19
$\langle x^2 J_{1\perp}\rangle/\sigma$	−3.01	−3.02	−2.97	−2.96	−2.98	−3.00	−3.00	−3.00
$\langle x q_\perp^2\rangle/\sigma$	2.43	2.32	2.25	2.22	2.20	2.18	2.16	2.14
$\langle x q_\parallel^2\rangle/\sigma$	2.61	2.45	2.35	2.30	2.27	2.24	2.20	2.17
$\langle x I_{2\perp}\rangle/\sigma$	0.35	0.39	0.40	0.41	0.43	0.44	0.46	0.46
$\langle J_{1\perp} q_\perp^2\rangle/\sigma$	−0.76	−0.80	−0.82	−0.83	−0.84	−0.86	−0.87	−0.89
$\langle J_{1\perp} q_\parallel^2\rangle/\sigma$	−1.50	−1.58	−1.61	−1.63	−1.67	−1.70	−1.72	−1.74

Table 6. Predicted σ's and three-point cumulants for ΛCDM power spectra in real space using PT.

R (Mpc h^{-1}):	8	16	24	32	40	48	56	64
σ	0.46	0.25	0.16	0.12	0.088	0.070	0.057	0.047
$\sigma_{1\perp}$	0.042	0.013	0.0058	0.0032	0.0020	0.0014	0.000 97	0.000 72
$\sigma_{1\parallel}$	0.030	0.0089	0.0041	0.0023	0.0014	0.000 96	0.000 67	0.000 51
$\sigma_{2\perp}$	0.0061	0.000 96	0.000 30	0.000 13	0.000 065	0.000 037	0.000 023	0.000 015
$\sigma_{2\parallel}$	0.0038	0.000 59	0.000 19	0.000 079	0.000 040	0.000 023	0.000 014	0.000 0093
$\langle x^3\rangle/\sigma$	3.70	3.52	3.43	3.38	3.34	3.31	3.28	3.25
$\langle x^2 J_{1\perp}\rangle/\sigma$	−3.21	−3.20	−3.15	−3.13	−3.16	−3.17	−3.17	−3.16
$\langle x q_\perp^2\rangle/\sigma$	2.51	2.39	2.32	2.28	2.25	2.23	2.21	2.19
$\langle x q_\parallel^2\rangle/\sigma$	2.51	2.39	2.32	2.28	2.25	2.23	2.21	2.19
$\langle x J_{2\perp}\rangle/\sigma$	0.38	0.42	0.43	0.45	0.46	0.48	0.49	0.50
$\langle J_{1\perp} q_\perp^2\rangle/\sigma$	−0.79	−0.83	−0.85	−0.86	−0.88	−0.90	−0.91	−0.92
$\langle J_{1\perp} q_\parallel^2\rangle/\sigma$	−1.58	−1.66	−1.70	−1.73	−1.76	−1.79	−1.82	−1.85

Now the integration of equation (93) over $\mu^2 k^2 (= (\boldsymbol{k}_1 + \boldsymbol{k}_2)\cdot\hat{\boldsymbol{z}})$ is straightforward; the integration over R^2 is also straightforward and involves no constant of integration. Cumulants integrated numerically on k_1, k_2 but analytically on the angles as described above are summed up in Table 4. Here, all variables are rescaled by their respective variance so that for instance $x = \delta/\sigma$, $q_\perp^2 = (\delta_1^2 + \delta_2^2)/\sigma_{1\perp}^2, \ldots$ Note that they are computed for $\Omega = 1$ (i.e. $f = 1$), $b_1 = 1$, $P(k) = k^n$ in the plane-parallel approximation and for a Gaussian filter. Skewness can be compared to Hivon et al. (1995) (where they do not assume a plane-parallel approximation). These cumulants are also computed for a Λ cold dark matter (ΛCDM) power spectrum ($\Omega_\mathrm{m} = 0.27$, $\Omega_\Lambda = 0.73$, $h = 0.7$) smoothed over different scales in Table 5 (redshift space) and Table 6 (real space).

In the regime where standard PT holds in redshift space,[10] we therefore may assume that all cumulants entering equations (65)–(69) can be *predicted* by a given standard cosmological model, while the amplitude of the NG correction scales like σ. For ΛCDM cosmology, Table 7 provides the predictions for the combinations of cumulants entering 3D Euler characteristic at first NG order, as a function of the smoothing length in real and in redshift space. The dependence on the linear bias b_1 of these particular combinations of cumulants can also be computed: for a ΛCDM power spectrum smoothed e.g. over 32 Mpc h^{-1}, it is found that $\tilde{S}^{(1)}_{\chi_{3\mathrm{D}},z}/\tilde{S}^{(1)}_{\chi_{3\mathrm{D}},r}$ is slightly varying around 0.8 while $\tilde{S}^{(-1)}_{\chi_{3\mathrm{D}},z}/\tilde{S}^{(-1)}_{\chi_{3\mathrm{D}},r}$ behaves approximately like $1.285 - 0.36/b_1$ for $1 < b_1 < 2.5$. Conversely, one may parametrize these cumulants while exploring alternative theories of gravity and attempt to fit these cumulants using geometric probes.

[10] Extensions of PT in redshift space using the streaming model (see Scoccimarro 2004; Taruya, Nishimichi & Saito 2010, for instance) would allow us to extend the validity of the predicted cumulants to smaller scales, but stand beyond the scope of this paper.

Table 7. Comparison of the reduced coefficients in front of the Hermite polynomials in the 3D Euler characteristic at first NG order for ΛCDM power spectra with different smoothings.

$R\,(\mathrm{Mpc}\,h^{-1})$:	8	16	24	32	40	48	56	64
$\tilde{S}^{(1)}$	0.075	0.062	0.044	0.033	0.036	0.039	0.038	0.035
$\tilde{S}_z^{(1)}$	0.061	0.053	0.037	0.028	0.033	0.038	0.039	0.037
$\tilde{S}^{(-1)}$	0.015	0.20	0.32	0.38	0.43	0.48	0.53	0.59
$\tilde{S}_z^{(-1)}$	−0.014	0.17	0.29	0.35	0.40	0.46	0.51	0.57

5.3.2 Constraining modified gravity with $\hat{\gamma}$ models

Over the last few years, many flavours of so-called modified gravity (MG) models have been presented (see, e.g., Clifton et al. 2012 for a review). In the context of the upcoming dark energy missions, it is of interest to understand how topological estimators also allow us to test such extensions of general relativity (GR). Following Bernardeau & Brax (2011), let us consider the so-called $\hat{\gamma}$ model as an illustrative example of how to test MG theories. This model parametrizes modifications of gravity through a change in the amplitude of Euler equation's source term. For this set of models, the generalization of equation (87) becomes parametric:

$$F_2^{\hat{\gamma}}(\boldsymbol{k}_1,\boldsymbol{k}_2)=\left(\frac{3\hat{\nu}_2(\hat{\gamma})}{4}-\frac{1}{2}\right)+\frac{1}{2}\frac{\boldsymbol{k}_1.\boldsymbol{k}_2}{k_1^2}+\frac{1}{2}\frac{\boldsymbol{k}_1.\boldsymbol{k}_2}{k_2^2}+\left(\frac{3}{2}-\frac{3\hat{\nu}_2(\hat{\gamma})}{4}\right)\frac{(\boldsymbol{k}_1.\boldsymbol{k}_2)^2}{k_1^2k_2^2}\,, \tag{94}$$

$$G_2^{\hat{\gamma}}(\boldsymbol{k}_1,\boldsymbol{k}_2)=\left(\frac{3\hat{\mu}_2(\hat{\gamma})}{4}-\frac{1}{2}\right)+\frac{1}{2}\frac{\boldsymbol{k}_1.\boldsymbol{k}_2}{k_1^2}+\frac{1}{2}\frac{\boldsymbol{k}_1.\boldsymbol{k}_2}{k_2^2}+\left(\frac{3}{2}-\frac{3\hat{\mu}_2(\hat{\gamma})}{4}\right)\frac{(\boldsymbol{k}_1.\boldsymbol{k}_2)^2}{k_1^2k_2^2}\,, \tag{95}$$

where $\nu_2(\hat{\gamma})$ and $\mu_2(\hat{\gamma})$ are given by

$$\hat{\nu}_2(\hat{\gamma})=\nu_2^{\mathrm{GR}}-\frac{10}{273}\left(\hat{\gamma}-\gamma^{\mathrm{GR}}\right)(1-\Omega_{\mathrm{m}})\Omega_{\mathrm{m}}^{\gamma^{\mathrm{GR}}-1},\quad \hat{\mu}_2(\hat{\gamma})=\mu_2^{\mathrm{GR}}-\frac{50}{273}\left(\hat{\gamma}-\gamma^{\mathrm{GR}}\right)(1-\Omega_{\mathrm{m}})\Omega_{\mathrm{m}}^{\gamma^{\mathrm{GR}}-1}\,, \tag{96}$$

with

$$\nu_2^{\mathrm{GR}}=\frac{4}{3}+\frac{2}{7}\Omega_{\mathrm{m}}^{-1/143},\quad \mu_2^{\mathrm{GR}}=-\frac{4}{21}+\frac{10}{7}\Omega_{\mathrm{m}}^{-1/143},\quad \gamma_{\mathrm{GR}}=\frac{6}{11}\,. \tag{97}$$

Then

$$Z_2^{\hat{\gamma}}(\boldsymbol{k}_1,\boldsymbol{k}_2)=b_1F_2^{\hat{\gamma}}(\boldsymbol{k}_1,\boldsymbol{k}_2)+f\mu^2G_2^{\hat{\gamma}}(\boldsymbol{k}_1,\boldsymbol{k}_2)+\frac{f\mu k}{2}\left[\frac{\mu_1}{k_1}\left(b_1+f\mu_2^2\right)+\frac{\mu_2}{k_2}\left(b_1+f\mu_1^2\right)\right]+\frac{b_2}{2}\,, \tag{98}$$

with $f=\Omega_{\mathrm{m}}^{\hat{\gamma}}$ and $F_2^{\hat{\gamma}}$, $G_2^{\hat{\gamma}}$ given by equation (94). Third-order cumulants, e.g. $\tilde{S}^{(1)}_{\chi 3\mathrm{D},z}$, in equations (75) and (76) become functions of $\hat{\gamma}$ and can now be fitted for $\hat{\gamma}$ to measured departure to Gaussianity in dark energy surveys. We computed numerically these cumulants for ΛCDM power spectra ($\Omega_{\mathrm{m}}=0.27$, $h_0=70$, $f=\Omega_{\mathrm{m}}^{\hat{\gamma}}$) with different $\hat{\gamma}$. For instance, the values of the observables, $\sigma\tilde{S}^{(1)}_{\chi 3\mathrm{D},z}$ and $\sigma\tilde{S}^{(-1)}_{\chi 3\mathrm{D},z}$ (that can be accessed by measuring the 3D Euler characteristic) are typically enhanced by a factor of 1 per cent when $\hat{\gamma}$ varies from 0.55 (GR) to 0.67 (DGP models). More generally, any parametrization of MG (illustrated here on $\hat{\gamma}$ models) can be implemented in this framework. A measurement of Minkowski functionals then leads to constraints on the parameters (e.g. $\hat{\gamma}$) of the theory.

5.4 Illustration on ΛCDM simulations: DM and halo catalogues

Let us now illustrate our statistics on the HORIZON 4π *N*-body simulation (Teyssier et al. 2009) which contains 4096^3 DM particles distributed in a 2 h^{-1}Gpc periodic box to validate our number count prediction in a more realistic framework. This simulation is characterized by the following ΛCDM cosmology: $\Omega_{\mathrm{m}}=0.24$, $\Omega_\Lambda=0.76$, $n_{\mathrm{s}}=0.958$, $H_0=73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\sigma_8=0.77$ within one standard deviation of *Wilkinson Microwave Anisotropy Probe* 3 results (Spergel et al. 2003). These initial conditions were evolved non-linearly down to redshift zero using the AMR code RAMSES (Teyssier 2002), on a 4096^3 grid. The motion of the particles was followed with a multigrid particle-mesh Poisson solver using a cloud-in-cell interpolation algorithm to assign these particles to the grid (the refinement strategy of 40 particles as a threshold for refinement allowed us to reach a constant physical resolution of 10 kpc; see the above-mentioned two references).

5.4.1 Minkowski functionals and extrema counts for DM

A measurement of the three-point cumulants in the 4π simulation is displayed in Table 8. It shows that redshift distortion has a small impact on three-point cumulants. We thus expect 3D Minkowski functionals to be weakly affected by redshift-space distortion. However, let us keep in mind that even if this difference is small, it should be of great importance to model it properly in the context of high-precision cosmology, especially since in 2D slices, the effect should be boosted.

Table 8. Measured cumulants in redshift space (real space) for different smoothings in the HORIZON 4π simulation. Note that the ratio of the field dispersion in redshift space versus real space at large scales is ≈1.17 which corresponds to $\beta \approx 0.5$ as expected.

R (Mpc h^{-1}):	16	23	32	45	64
σ	0.28(0.25)	0.20(0.17)	0.14(0.12)	0.090(0.077)	0.057(0.049)
$\sigma_{1\|\|}$	0.044(0.035)	0.023(0.018)	0.012(0.0092)	0.0057(0.0045)	0.0027(0.0021)
$\sigma_{1\perp}$	0.054(0.050)	0.028(0.026)	0.014(0.013)	0.0068(0.0062)	0.0032(0.0029)
$\sigma_{2\|\|}$	0.012(0.0094)	0.0045(0.0035)	0.0017(0.0013)	0.000 59(0.000 45)	0.000 21(0.000 16)
$\sigma_{2\perp}$	0.016(0.015)	0.0061(0.0057)	0.0022(0.0021)	0.000 76(0.000 72)	0.000 26(0.000 24)
σ_Q	0.0089(0.0076)	0.0034(0.0029)	0.0012(0.0010)	0.000 43(0.000 36)	0.000 15(0.000 12)
$\langle x^3\rangle/\sigma$	3.36(3.49)	3.27(3.35)	3.24(3.28)	3.38(3.41)	4.03(4.07)
$\langle x^2 J_{1\perp}\rangle/\sigma$	−2.92(−3.19)	−2.84(−3.07)	−2.75(−2.95)	−2.76(−2.95)	−3.00(−3.20)
$\langle x q_\perp^2\rangle/\sigma$	2.27(2.40)	2.28(2.27)	2.09(2.16)	2.05(2.12)	2.19(2.26)
$\langle x q_\parallel^2\rangle/\sigma$	2.33(2.41)	2.27(2.30)	2.24(2.24)	2.24(2.21)	2.52(2.53)
$\langle x I_{2\perp}\rangle/\sigma$	0.39(0.43)	0.39(0.42)	0.37(0.41)	0.37(0.40)	0.42(0.46)
$\langle J_{1\perp} q_\perp^2\rangle/\sigma$	−0.81(−0.87)	−0.79(−0.84)	−0.76(−0.80)	−0.72(−0.76)	−0.81(−0.86)
$\langle J_{1\perp} q_\parallel^2\rangle/\sigma$	−1.57(−1.75)	−1.55(−1.68)	−1.53(−1.64)	−1.51(−1.60)	−1.53(−1.64)
β_σ	0.24(0.0066)	0.26(0.014)	0.28(0.030)	0.30(0.058)	0.33(0.091)

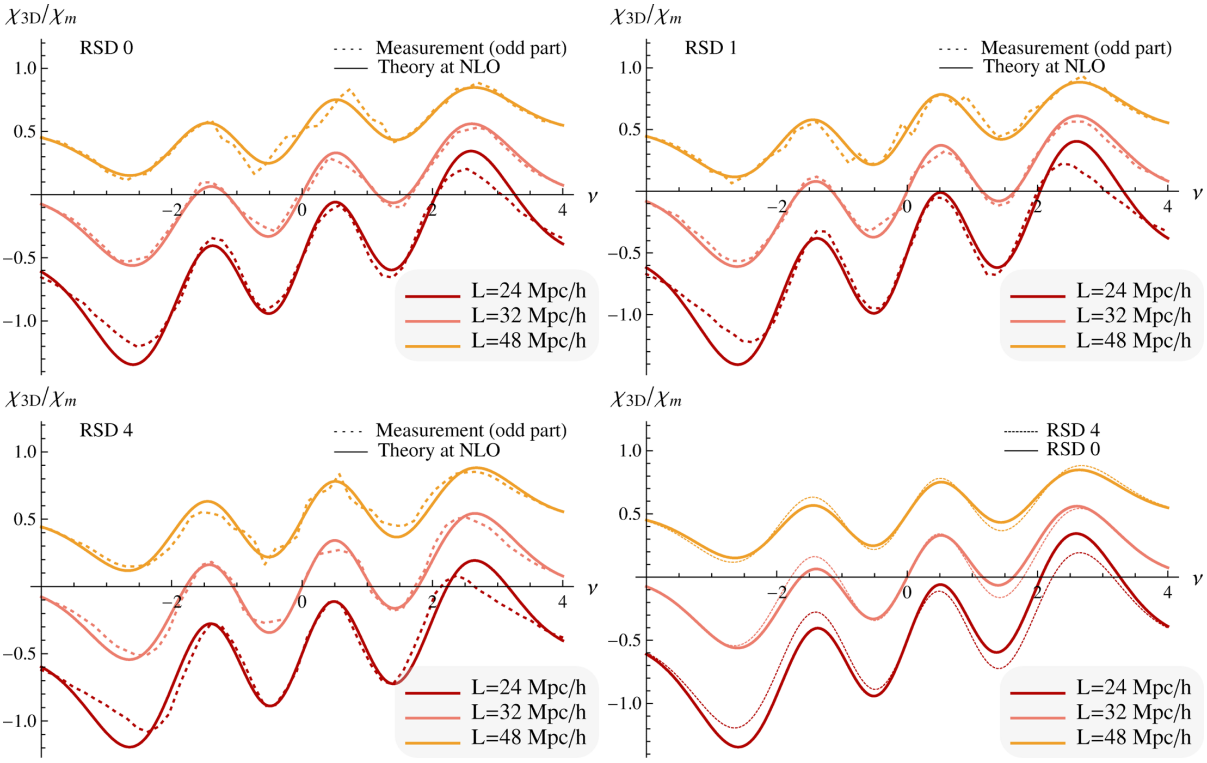


Figure 7. Top-left panel: first NG correction of the 3D Euler characteristic as a function of smoothing (as labelled) measured (dotted lines) and predicted (plain line) in HORIZON 4π simulation in real space. Each curve has been normalized by the maximum of the Gaussian component of the Euler characteristic. Those corresponding to different smoothing lengths have been shifted for clarity. Top-right panel: same as the top-left panel, but in redshift space. Bottom-left panel: same as the top-right panel, but redshift displacements were boosted by a factor of 4. Bottom-right panel: theoretical prediction in real space and boosted redshift space. For these ranges of smoothing, the theory predicts well the Euler characteristic to first order in non-Gaussianity, in particular for low-intermediate thresholds. The difference in the Euler characteristic introduced by redshift distortion (even boosted by a factor of 4) is rather small.

Fig. 7 illustrates this for the 3D Euler characteristic. First, note that the theory mimics very well the measurement. This figure also shows that the difference between real and redshift space is indeed small for the first correction from Gaussianity. But in 2D, this difference increases as seen in Fig. 8 which shows how the correction from Gaussianity depends on the angle between the LOS and the slice on which the 2D Euler characteristic is computed.

5.4.2 Minkowski functionals for halo catalogues

The friend-of-friend algorithm (Huchra & Geller 1982) was used over 18^3 overlapping subsets of the simulation with a linking length of 0.2 times the mean interparticle distance to define DM haloes. In the present work, we only consider haloes with more than 40 particles, which

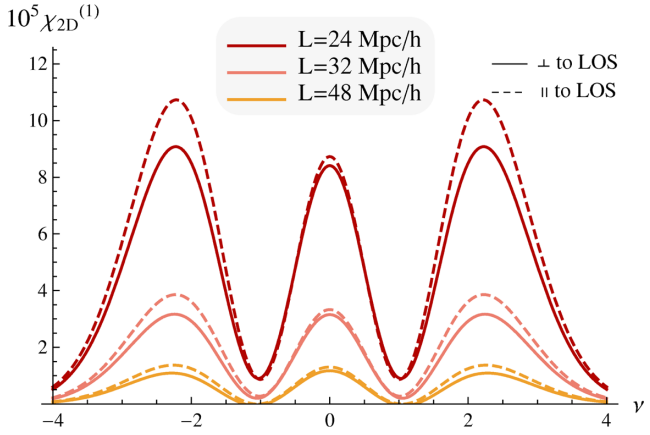


Figure 8. Prediction for the 2D Euler characteristic in redshift space along (dashed lines) and perpendicular (solid lines) to the LOS with cumulants measured in the HORIZON 4π simulation. There is a clear dependence on the angle between the slice and the LOS, especially for smaller smoothing.

corresponds to a minimum halo mass of $3 \times 10^{11}\,\mathrm{M}_\odot$ (the particle mass is $7.7 \times 10^9\,\mathrm{M}_\odot$). The mass dynamical range of this simulation spans about five decades. Overall, the catalogue contains 43 million dark haloes. The density of dark haloes in real space and redshift space was re-sampled over a 512^3 grid with box size 2000 Mpc h^{-1} (in redshift space, after shifting the z-ordinate of the halo by its velocity along that direction divided by the Hubble constant, H_0) and smoothed with a Gaussian filter of width $\sigma = 16, 24, \ldots$ up to 94 Mpc h^{-1}. For each cube in real and redshift space, the number density of extrema is computed by a local quadratic fit to the function profile with control over the double counting of the neighbouring extrema [see Pogosyan et al. (2011) for the technique description and Colombi, Pogosyan & Souradeep (2000) for a first application]. The relevant 35 cumulants are computed via fast Fourier transform. Fig. 9 displays the corresponding predicted (solid line) and measured (dashed line) number count difference in real space (left-hand panel) and redshift space (right-hand panel). Note that redshift- and real-space extrema counts are almost indistinguishable. Indeed, for this biased population, b_1 is large ($\sim$2) and f/b_1 is small enough.

5.5 A cosmic fiducial experiment: measuring β and σ via slicing along and perpendicular to the LOS

As mentioned in Section 4, the angle dependence of the 2D Minkowski functionals allows us to probe β. Indeed, equations (43), (54) and (60) display a functional dependence on $\theta_{\mathcal{S}/\mathcal{L}}$. At Gaussian order first, it appears that

$$\chi_{2D}^{(0)}(\nu, \theta_{\mathcal{S}}) \propto \sqrt{1 - \beta_\sigma \sin^2 \theta_{\mathcal{S}}}\,, \tag{99}$$

so that measuring the Gaussian part of the 2D Euler characteristic in different slices (with different orientations relative to the LOS) can give access to $\beta_\sigma = 1 - \sigma_{1\perp}^2/2\sigma_{1\parallel}^2 = \frac{4}{5}\beta + \mathcal{O}(\beta^2)$. For example, in the most favourable case of $\theta_1 = \pi/2$ and $\theta_2 = 0$, one have direct access to

$$\frac{\chi_{2D}^{(0)}(\nu, \theta_1)}{\chi_{2D}^{(0)}(\nu, \theta_2)} = \sqrt{\frac{1 - \beta_\sigma \sin^2 \theta_1}{1 - \beta_\sigma \sin^2 \theta_2}} = \sqrt{1 - \beta_\sigma} = 1 - \frac{2}{5}\beta + \mathcal{O}(\beta^2)\,. \tag{100}$$

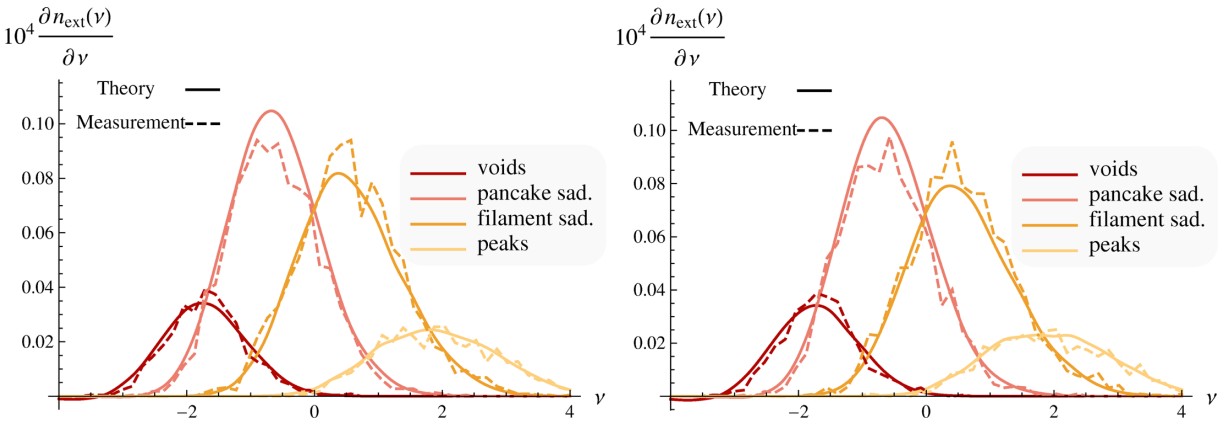


Figure 9. Left-hand panel: the distribution of 3D extrema in real space as a function of threshold ν in real space, as predicted (solid) and measured (dashed) in DM halo catalogue of the HORIZON 4π simulation smoothed over 48 Mpc h^{-1}. Right-hand panel: same as the left-hand panel but in redshift space.

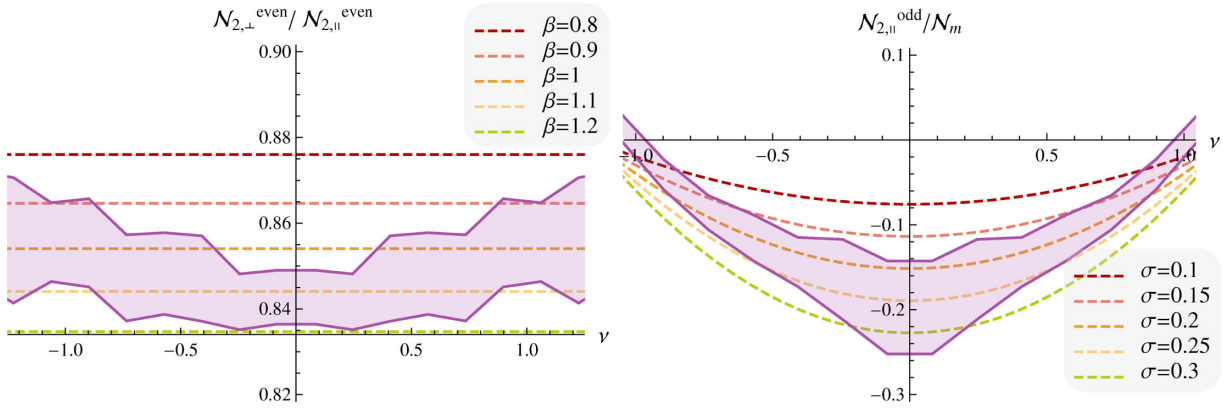


Figure 10. Reconstructed β from the Gaussian part of $\mathcal{N}_2$ (left-hand panel) and σ from its NG correction (right-hand panel).

Contour-crossing statistics have the same angle dependence as the 2D Euler characteristic at Gaussian order. The $\mathcal{N}_2$ statistics (area of isocontours) leads in turn to

$$\frac{\mathcal{N}_2^{(0)}(\nu,\theta_1)}{\mathcal{N}_2^{(0)}(\nu,\theta_2)} = \frac{E(\tilde{\beta}_\sigma(\theta_1))\sqrt{1-\tilde{\beta}_\sigma(\theta_2)}}{E(\tilde{\beta}_\sigma(\theta_2))\sqrt{1-\tilde{\beta}_\sigma(\theta_1)}} = \frac{\pi}{2}\frac{\sqrt{1-\beta_\sigma}}{E(\beta_\sigma)} = 1-\frac{1}{5}\beta+\mathcal{O}(\beta^2)\,. \tag{101}$$

Note that beyond a simple overall amplitude effect (which is enough to measure β alone), this angle dependence also arises in the first NG correction as plotted in Fig. 5.

If the value of β is important on its own (e.g. to study the bias or to test modifications of gravity), it is also of prime importance to measure $D(z)$ as it allows us to map the dispersion in redshift space, σ, into its value in real space, $\sigma_0 \propto D(z)$. One way to proceed is to use the Gaussian term to put constraints on β and then the NG correction for σ (the amplitude of which is not probed by the Gaussian part). Indeed, following Gay et al. (2012), the theory presented in Section 4 should allow us to measure the dispersion of the field in redshift space σ from the amplitude of the departure from non-Gaussianity. As equation (101) demonstrates, the comparison of the 2D Minkowski functionals in planes parallel and perpendicular to the LOS allows us to measure independently $\beta = f/b$. Hence, Minkowski functionals in redshift space yield a geometric estimate of the real-space field dispersion, σ_0, via

$$\sigma_0 = \sigma/\sqrt{1+2\beta/3+\beta^2/5}. \tag{102}$$

Let us apply this scheme to measure β and σ_0 in a fiducial experiment. Minkowski functionals are measured from our set of 19 scale-invariant ($n=-1$) 256^3 DM simulations smoothed over 15 pixels, corresponding to $\sigma_0 = 0.18$ and displayed in Fig. 4. A first step is to use the Gaussian term of the 1D and 2D statistics, while varying the angle of the slices to constrain β using equations (100) and (101). For that purpose, we extract the even part of $\mathcal{N}_2$ (to get rid of odd parity effects arising from the first NLO correction) and restrict ourselves to the intermediate domain $-1.2 < \nu < 1.2$ where the Gaussian term is dominant. The resulting constraints on β are found to be $\hat{\beta} = 1.04 \pm 0.05$, and illustrated in Fig. 10 (left-hand panel). The same analysis on the other 1D ($\mathcal{N}_1$) and 2D statistics (2D genus) leads to similar constraints.

The next step is to fit the first NG correction of each statistics with PT predictions in order to constrain σ. The predictions at first order are expressed as a function of the contrast ν and σ only using Table 4. The value of the free parameter σ in the model is then constrained by fitting the odd part of the data (which is dominated by the first NG correction for intermediate contrasts). The result for $\mathcal{N}_2$ is shown in Fig. 10 (right-hand panel) and yields $\hat{\sigma} = 0.22 \pm 0.08$. The other statistics give similar results e.g. for the 3D Euler characteristic $\hat{\sigma} = 0.26 \pm 0.06$. Altogether, using σ as measured by a 3D probe, χ_{3D}, and β by a 2D statistics, $\mathcal{N}_2$, we finally get $\hat{\sigma}_0 = 0.18 \pm 0.04$, which is fully consistent with the underlying dispersion in our mocks. The accuracy on the measurement of β and $D(z)$ through σ_0 can naively be scaled to the expected accuracy for a *Euclid*-like survey (assuming one quarter of the sky is observed) leading to a relative 0.3 per cent precision on β and 1.5 per cent on $D(z)$ at redshift zero. See also Gay et al. (2012), which translates this accuracy in terms of estimates for the dark energy parameters w_0, w_a.

It is worth noting that in this simple fiducial experiment, several assumptions were made: first, we assumed we knew the contrast, ν, while in realistic surveys, the accessible quantity is ν_f (see Section 4.5.1); for the simulation, the cosmology is not ΛCDM but Einstein–de Sitter with a scale-invariant initial power spectrum; finally the error on the estimated σ depends on the accuracy of the theory used to predict the cumulants (standard PT here, which is known to perform somewhat poorly in redshift space). No account of masking, redshift evolution of the S/N ratio or finite survey volume, nor comparison with other dark energy probes was attempted. Carrying out the road map sketched in this section while addressing these issues should be one of the target of the upcoming surveys that have been planned specifically to probe dark energy, either from ground-based facilities (e.g. BigBOSS, VST-KIDS, DES, Pan-STARRS, LSST[11]) or from space-based observatories [*Euclid* (Laureijs et al. 2011), SNAP and JDEM[12]].

[11] http://bigboss.lbl.gov, http://wtww.astro-wise.org/projects/KIDS, https://www.darkenergysurvey.org, http://pan-starrs.ifa.hawaii.edu, http://www.lsst.org
[12] http://sci.esa.int/euclid, http://snap.lbl.gov, http://jdem.lbl.gov

6 CONCLUSION

This paper has placed on a firm footing the statistical analysis of topological sets in anisotropic spaces. Specifically, it has presented extensions of Matsubara (1996) and Gay et al. (2012) in two directions: it has accounted for anisotropic fields and non-Gaussianity for all Minkowski functionals. The main results of this work are as follows: (i) the anisotropic JPDF in 2D and 3D: a new building block for redshift-space analysis; (ii) the analytical Euler characteristic at all orders in 2D/3D (and therefore the rare event limit for extrema counts); (iii) other Minkowski functionals and extrema counts to first order in NG correction in anisotropic space; (iv) extension of PT for the relevant cumulants. The theory presented in Sections 3 and 4 makes no assumption about the origin of the anisotropy and could therefore be implemented in contexts beyond astrophysics.

In the context of cosmology, the theoretical expectations were robustly derived both in invariant and field variables, and checked against Monte Carlo simulations of $f_{\rm nl}$ models, scale-invariant DM simulations and ΛCDM simulations (both for DM and halo catalogues). It was shown how to use these predictions to measure σ_0 and β as a function of redshift. The implication of the invariance of Minkowski functionals and extrema counts versus monotonic mapping on the relevant combinations of cumulants (i.e. projections of the bispectrum) was made explicit. Its relevance for MG probes was discussed. The implementation on DM haloes of ΛCDM simulations for biased populations with $M > 10^{11}{\rm M}_\odot$ allowed us to quantify and formalize the (weak) effect of redshift distortion on the 3D geometric descriptors of the field. This weak sensitivity reflects the robustness of topological estimates. In contrast, the comparison of 2D slices perpendicular and parallel to the LOS should allow us to also measure $\sigma(z)$ and correct it for $\Omega_{\rm m}^{\gamma}/b_1$ as demonstrated in a fiducial experiment. At scales above $\sim 50\,{\rm Mpc}\,h^{-1}$, it has been found that standard PT allows us to predict the theoretical cumulants at the level of accuracy required to match the measured cumulants from simulations within a range of contrast $\nu \in [-2, 2]$. As Minkowski functionals are configuration space probes, we can expect a better convergence of (extended) PT for these contrasts which are less sensitive to the dynamics of very non-linear regions.

Improvements beyond the scope of this paper include (i) improving PT in redshift space while implementing variations of the streaming model (see Scoccimarro 2004; Taruya et al. 2010, for instance) and/or anisotropic smoothing; (ii) departing from the plane-parallel approximation while constructing a full-sky prescription for NG Minkowski functionals of realistic catalogues; (iii) extending the prediction to the statistics of the skeleton and walls; (iv) propagating to cosmic parameter estimation, the residual misfits; (v) extending the prediction of the JPDF to *N*-point statistics for non-local analysis (e.g. void size and non-linear *N*-point peak statistics); (vi) exploring alternative expansion to the GC's; (vii) deriving the statistics of *errors* on one-point statistics such as those presented in this paper; (viii) implementing the relevant theory on realistic mocks and demonstrating pros and cons of geometrical probes (e.g. in the presence of masks), and contrast those to existing dark energy probes (lensing, SN1a, etc.).

In the context of upcoming 3D spectroscopic surveys such as *Euclid*, the statistical analysis of the geometry of our redshift-distorted Universe will allow us to robustly measure weighted moments of the multispectra as a function of redshift, and henceforth quantify the cosmic evolution of the equation of state of dark energy and possible departure from GR.

ACKNOWLEDGEMENTS

We thank S. Colombi, T. Nishimichi and S. Prunet for useful comments during the course of this project, and Lena for her hospitality when part of this work was completed. DP thanks the LABEX 'Institut de Lagrange de Paris' and the 'Programme National de Cosmologie' for funding. DP's research was supported in part by the National Science Foundation under Grant No. NSF PHY11-25915. SC and CP thank the University of Alberta for funding. This research is part of the Horizon-UK project and the ANR Cosmo@NLO (ANR-12-BS05-0002) and Spin(E). Special thanks to T. Sousbie for his help in producing the friend-of-friend catalogue for the HORIZON 4π simulation and to our collaborators of the Horizon project (www.projet-horizon.fr) for helping us produce the simulation. We warmly also thank S. Prunet and S. Rouberol for running the `Horizon` cluster for us. Let us finally thank D. Munro for freely distributing his YORICK programming language and opengl interface (available at yorick.sourceforge.net) and the community of mathematica.stackexchange.com for their help.

REFERENCES

Adler R. J., 1981, The Geometry of Random Fields. Wiley, Chichester
Amendola L., 1996, MNRAS, 283, 983
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bernardeau F., Brax P., 2011, J. Cosmol. Astropart. Phys., 6, 19
Bernardeau F., Kofman L., 1995, ApJ, 443, 479
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Blinnikov S., Moessner R., 1998, A&AS, 130, 193
Bond J. R., Efstathiou G., 1987, MNRAS, 226, 655
Chambers J. M., 1967, Biometrika, 54, 367
Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Phys. Rep., 513, 1
Coles P., 1988, MNRAS, 234, 509
Colombi S., Pogosyan D., Souradeep T., 2000, Phys. Rev. Lett., 85, 5515
Cramér H., 1946, Mathematical Methods of Statistics. Princeton Univ. Press, Princeton, NJ
Desjacques V., Sheth R. K., 2010, Phys. Rev. D, 81, 023526
Doroshkevich A. G., 1970, Astrofizika, 6, 581
Gay C., Pichon C., Pogosyan D., 2012, Phys. Rev. D, 85, 023011

Gaztañaga E., Cabré A., Castander F., Crocce M., Fosalba P., 2009, MNRAS, 399, 801
Gott J. R., III, 1988, PASP, 100, 1307
Gott J. R., III, Weinberg D. H., Melott A. L., 1987, ApJ, 319, 1
Gott J. R., III et al., 1989, ApJ, 340, 625
Gott J. R., III, Park C., Juszkiewicz R., Bies W. E., Bennett D. P., Bouchet F. R., Stebbins A., 1990, ApJ, 352, 1
Gott J. R. I., Colley W. N., Park C.-G., Park C., Mugnolo C., 2007, MNRAS, 377, 1668
Hadwiger H., 1957, Vorlesungen über Inhalt, Oberfläche und Isoperimetrie. Springer
Hamilton A. J. S., Gott J. R., III, Weinberg D., 1986, ApJ, 309, 1
Heyvaerts J., Pichon C., Prunet S., Thiébaut J., 2013, MNRAS, 430, 3320
Hikage C. et al., 2002, PASJ, 54, 707
Hikage C. et al., 2003, PASJ, 55, 911
Hivon E., Bouchet F. R., Colombi S., Juszkiewicz R., 1995, A&A, 298, 643
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Jackson J. C., 1972, MNRAS, 156, 1P
Jost J., 2008, Riemannian Geometry and Geometric Analysis, 5th edn. Springer, Berlin
Juszkiewicz R., Weinberg D. H., Amsterdamski P., Chodorowski M., Bouchet F., 1995, ApJ, 442, 39
Kaiser N., 1984, ApJ, 284, L9
Kaiser N., 1987, MNRAS, 227, 1
Kendall M. G., Stuart A., 1958, The Advanced Theory of Statistics. Vol. 1, Griffin, London
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lazarian A., Pogosyan D., 2000, ApJ, 537, 720
Matsubara T., 1994, ApJ, 434, L43
Matsubara T., 1996, ApJ, 457, 13
Matsubara T., 2003, ApJ, 584, 1
Matsubara T., 2010, Phys. Rev. D, 81, 083505
Matsubara T., Suto Y., 1995, ApJ, 460, 51
Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697
Melott A. L., Weinberg D. H., Gott J. R., III, 1988, ApJ, 328, 50
Melott A. L., Cohen A. P., Hamilton A. J. S., Gott J. R., III, Weinberg D. H., 1989, ApJ, 345, 618
Nishimichi T., Taruya A., Koyama K., Sabiu C., 2010, J. Cosmol. Astropart. Phys., 7, 2
Novikov D., Colombi S., Doré O., 2003, MNRAS, 366, 15
Park C. et al., 2005, ApJ, 633, 11
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C., Szalay A. S., 2007, MNRAS, 381, 1053
Planck Collaboration, 2013, preprint (arXiv:1303.5083)
Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F., Sousbie T., Colombi S., 2009a, MNRAS, 396, 673
Pogosyan D., Gay C., Pichon C., 2009b, Phys. Rev. D, 80, 081301
Pogosyan D., Pichon C., Gay C., 2011, Phys. Rev. D, 84, 083510
Rice S. O., 1944, Bell Syst. Tech. J., 23, 282
Rice S. O., 1945, Bell Syst. Tech. J., 24, 46
Ryden B. S., 1988, ApJ, 333, L41
Ryden B. S., Melott A. L., Craig D. A., Gott J. R., III, Weinberg D. H., Scherrer R. J., Bhavsar S. P., Miller J. M., 1989, ApJ, 340, 647
Sargent W. L. W., Turner E. L., 1977, ApJ, 212, L3
Scaramella R., Vittorio N., 1991, ApJ, 375, 439
Scoccimarro R., 2004, Phys. Rev. D, 70, 083007
Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, ApJ, 496, 586
Scoccimarro R., Couchman H. M. P., Frieman J. A., 1999, ApJ, 517, 531
Seto N., 2000, ApJ, 538, 11
Spergel D. N. et al., 2003, ApJS, 148, 175
Taruya A., Nishimichi T., Saito S., 2010, Phys. Rev. D, 82, 063522
Teyssier R., 2002, A&A, 385, 337
Teyssier R. et al., 2009, A&A, 497, 335
Verde L., Heavens A. F., Matarrese S., Moscardini L., 1998, MNRAS, 300, 747
Wang X., Chen X., Park C., 2012, ApJ, 747, 48
Weinberg D. H., Gott J. R., III, Melott A. L., 1987, ApJ, 321, 2
Zunckel C., Gott J. R., Lunnan R., 2011, MNRAS, 412, 1401

APPENDIX A: SPECTRAL PARAMETERS AND CUMULANTS

A1 Relations with the parametrization of Matsubara (1996)

At linear order, the redshift distortion, equation (83), reduces to the Kaiser (1987) result in equation (12). Matsubara (1996) defines the dispersions and angular moments,

$$\sigma_j^2(R) \equiv \frac{1}{2\pi^2}\int \mathrm{d}k\, k^{2+2j} P^{(r)}(k)\, W^2(kR)\,, \quad C_j(f/b) = \frac{1}{2}\int_{-1}^{1} \mathrm{d}\mu\, \mu^{2j}(1+fb^{-1}\mu^2)^2\,, \tag{A1}$$

and uses the normalized quantities,

$$\alpha \equiv \frac{\delta_R^{(s)}}{\sigma_0\sqrt{C_0}}, \quad \beta_i \equiv \frac{\partial_i \delta_R^{(s)}}{\sigma_0\sqrt{C_0}} \quad \text{and} \quad \omega_{ij} \equiv \frac{\partial_i \partial_j \delta_R^{(s)}}{\sigma_0\sqrt{C_0}}, \tag{A2}$$

to describe the smoothed density contrast in redshift space $\delta_R^{(s)}$ and its derivatives. Our variables relate to these as

$$x = \alpha, \quad x_I = \frac{\sqrt{C_0}\sigma_0}{\sigma_{1\perp}}\beta_I, \quad x_3 = \frac{\sqrt{C_0}\sigma_0}{\sigma_{1\parallel}}\beta_3, \quad x_{IJ} = \frac{\sqrt{C_0}\sigma_0}{\sigma_{2\perp}}\omega_{IJ} \text{ and } x_{33} = \frac{\sqrt{C_0}\sigma_0}{\sigma_{2\parallel}}\omega_{33}, \tag{A3}$$

where $I, J \in \{1, 2\}$, so that

$$\sigma^2 = C_0\sigma_0^2, \quad \sigma_{1\perp}^2 = (C_0 - C_1)\sigma_1^2, \quad \sigma_{1\parallel}^2 = C_1\sigma_1^2, \quad \sigma_{2\perp}^2 = (C_0 - 2C_1 + C_2)\sigma_2^2, \quad \sigma_{2\parallel}^2 = C_2\sigma_2^2, \tag{A4}$$

with

$$C_0 = 1 + \frac{2f}{3b} + \frac{f^2}{5b^2}, \quad C_1 = \frac{1}{3} + \frac{2f}{5b} + \frac{f^2}{7b^2}, \quad C_2 = \frac{1}{5} + \frac{2f}{7b} + \frac{f^2}{9b^2}\,. \tag{A5}$$

In terms of these variables, both formulations are in exact agreement. In the isotropic Gaussian limit (i.e. the linear regime with no redshift correction), $C_0 = 1$, $C_1 = 1/3$ and $C_2 = 1/5$.

A2 Inter-relations between the moments of random fields and simplification of Gay et al. (2012) results

Not all cumulants of the field and its derivatives that appear in the GC expansion are independent. Relations between the cumulants can be established, for instance, by expressing statistical averages as averages over spatial coordinates using homogeneity of the statistics and utilizing integration by parts. The details of the results depend on the properties of the manifold; in this paper we limit ourselves to the theory in the flat infinite Euclidean space. As an example,

$$\left\langle x^2 x_{33}\right\rangle = \lim_{V\to\infty}\frac{1}{V}\int_V \mathrm{d}V x^2 x_{33} = -\gamma_\parallel \lim_{V\to\infty}\frac{2}{V}\int_V \mathrm{d}V x x_3^2 = -2\gamma_\parallel \left\langle x x_3^2\right\rangle\,, \tag{A6}$$

where the boundary term vanishes as volume V is taken in the infinite limit. A similar procedure shows that, e.g., $\langle x_1^2 x_{11}\rangle = 0$.

Proceeding in this manner one can establish for isotropic fields several useful general relations that involve the rotation invariants I_s of the Hessian matrix x_{ij}:[13]

$$\langle x^n I_1\rangle = -n\gamma \left\langle x^{n-1} q^2\right\rangle, \tag{A7}$$

$$\langle x^n I_2\rangle = -\frac{3}{4}n\gamma \left\langle x^{n-1} q^2 I_1\right\rangle - \frac{1}{4}n(n-1)\gamma^2 \left\langle x^{n-2} q^4\right\rangle, \tag{A8}$$

$$\langle x^n I_s\rangle = -\frac{1}{s}\sum_{k=1}^{\min(s,n)} \frac{s+k}{(2k)!!}\frac{n!}{(n-k)!}\gamma^k \left\langle x^{n-k} q^{2k} I_{s-k}\right\rangle, s >= 1, I_0 = 1. \tag{A9}$$

These relations are valid in any dimensions, for correspondingly defined I_s's,[14] and also hold for the cumulants as well as the moments.

A2.1 Isotropic 3D Euler characteristic

The relations (A7)–(A9) allow us to simplify the expression for the Euler characteristic given by Gay et al. (2012) for isotropic fields. Indeed, these authors found[15]

$$\chi_{3\mathrm{D}}^{\mathrm{iso}(0)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^2 R_\star^3}\frac{\sqrt{3}}{9}\gamma^3 H_2(\nu)\,, \tag{A10}$$

$$\chi_{3\mathrm{D}}^{\mathrm{iso}(1)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^2 R_\star^3}\frac{\sqrt{3}}{9}\left[\frac{1}{6}\gamma^3\left\langle x^3\right\rangle H_5(\nu) - \frac{3}{2}\left(\gamma^3\left\langle x q^2\right\rangle + \gamma^2\left\langle x^2 J_1\right\rangle\right) H_3(\nu) + 9\left(\frac{1}{2}\gamma^2\left\langle q^2 J_1\right\rangle + \gamma\left\langle x I_2\right\rangle\right) H_1(\nu)\right], \tag{A11}$$

[13] I_s are the coefficients of the characteristic polynomial of the matrix x_{ij}, namely in N dimensions I_1 is the trace, I_N is the determinant and for $1 < s < N$, I_s is the sum of the minors of order s of x_{ij}.

[14] These relations can be generalized to homogeneous spaces of constant curvature, e.g. 2-sphere, with additional terms that are proportional to the curvature of the space appearing during some integrations by parts.

[15] We shall take this opportunity to correct several unfortunate misprints in expressions for the second-order corrections $\chi^{(2)}$ as presented in Gay et al. (2012). In this text, equations (A12) and (A17) are corrected versions.

$$\chi_{3D}^{\mathrm{iso}(2)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^2 R_\star^3}\frac{\sqrt{3}}{9}\left[-\left(\frac{27}{2}\gamma\left\langle q^2 I_2\right\rangle_{\mathrm{c}} + 27\langle x I_3\rangle_{\mathrm{c}}\right)H_0(\nu) + \left(\frac{9}{8}\gamma^3\left\langle q^4\right\rangle_c + \frac{9}{2}\gamma^2\left\langle xq^2 J_1\right\rangle_{\mathrm{c}} + \frac{9}{2}\gamma\left\langle x^2 I_2\right\rangle_{\mathrm{c}}\right)H_2(\nu)\right.$$
$$- \left(\frac{3}{4}\gamma^3\left\langle x^2q^2\right\rangle_{\mathrm{c}} + \frac{1}{2}\gamma^2\langle x^3 J_1\rangle_{\mathrm{c}}\right)H_4(\nu) + \frac{1}{24}\gamma^3\langle x^4\rangle_{\mathrm{c}}H_6(\nu) + \left(\frac{135}{16}\gamma\langle q^2 J_1\rangle^2 + \frac{27}{2}\langle q^2 J_1\rangle\langle x I_2\rangle + \frac{81}{2}\langle xq^2\rangle\langle I_3\rangle\right)H_0(\nu)$$
$$- \left(\frac{45}{4}\gamma^2\langle xq^2\rangle\langle q^2 J_1\rangle + \frac{9}{2}\gamma\langle q^2 J_1\rangle\langle x^2 J_1\rangle + \frac{27}{2}\gamma\langle xq^2\rangle\langle x I_2\rangle + \frac{9}{2}\langle x^2 J_1\rangle\langle x I_2\rangle + \frac{9}{2}\langle x^3\rangle\langle I_3\rangle\right)H_2(\nu)$$
$$+ \left(\frac{15}{8}\gamma^3\langle xq^2\rangle^2 + \frac{3}{4}\gamma^2\langle x^3\rangle\langle q^2 J_1\rangle + \frac{9}{4}\gamma^2\langle xq^2\rangle\langle x^2 J_1\rangle + \frac{3}{4}\gamma\langle x^2 J_1\rangle^2 + \frac{3}{2}\gamma\langle x^3\rangle\langle x I_2\rangle\right)H_4(\nu)$$
$$\left. - \left(\frac{1}{4}\gamma^3\langle xq^2\rangle\langle x^3\rangle + \frac{1}{4}\gamma^2\langle x^3\rangle\langle x^2 J_1\rangle\right)H_6(\nu) + \frac{1}{72}\gamma^3\langle x^3\rangle^2 H_8(\nu)\right], \tag{A12}$$

where $R_\star = \sigma_1/\sigma_2$. Using equations (A7)–(A9), the equations (A10) and (A11) can be rewritten up to NLO as

$$\chi_{3D}^{\mathrm{iso}(0+1)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^2}\frac{\sigma_1^3}{\sigma^3}\frac{\sqrt{3}}{9}\left[H_2(\nu) + \frac{1}{3!}H_5(\nu)\langle x^3\rangle + \frac{3}{2}H_3(\nu)\langle xq^2\rangle - \frac{15}{2}H_1(\nu)\frac{1}{\gamma}\left\langle q^2 J_1\right\rangle\right], \tag{A13}$$

which is the form we quote in equation (38). The NNLO, $\mathcal{O}(\sigma^2)$, term can also be simplified from equation (A12) to a more compact form,

$$\chi_{3D}^{\mathrm{iso}(2)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^2}\frac{\sigma_1^3}{\sigma^3}\frac{\sqrt{3}}{9}\left[H_0(\nu)\left(\frac{9}{2}\gamma^{-2}\left\langle q^2 I_2\right\rangle_c - \frac{27}{16}\gamma^{-2}\langle J_1 q^2\rangle^2\right) + H_2(\nu)\left(-\frac{9}{8}\langle q^4\rangle_c - \frac{9}{4}\gamma^{-1}\left\langle xq^2 J_1\right\rangle_c + \frac{9}{8}\gamma^{-1}\langle xq^2\rangle\langle J_1 q^2\rangle\right)\right.$$
$$\left. + H_4(\nu)\left(\frac{3}{4}\left\langle x^2q^2\right\rangle_c + \frac{3}{8}\langle xq^2\rangle^2 - \frac{3}{8}\gamma^{-1}\langle x^3\rangle\left\langle J_1 q^2\right\rangle\right) + H_6(\nu)\left(\frac{1}{24}\langle x^4\rangle_c + \frac{1}{4}\langle x^3\rangle\langle xq^2\rangle\right) + H_8(\nu)\frac{1}{72}\langle x^3\rangle^2\right]. \tag{A14}$$

A2.2 Isotropic 2D Euler characteristic

The same procedure holds for the 2D Euler characteristic for which the expression given by Gay et al. (2012) in terms of a 2D isotropic field x

$$\chi_{2D}^{\mathrm{iso}(0)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}}\frac{1}{2R_\star^2}\gamma^2 H_1(\nu), \tag{A15}$$

$$\chi_{2D}^{\mathrm{iso}(1)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}}\frac{1}{2R_\star^2}\left[\frac{1}{3!}\gamma^2\langle x^3\rangle H_4(\nu) - \left(\gamma^2\langle xq^2\rangle + \gamma\langle x^2 J_1\rangle\right)H_2(\nu) + \left(2\gamma\langle q^2 J_1\rangle + 4\langle x I_2\rangle\right)\right], \tag{A16}$$

$$\chi_{2D}^{\mathrm{iso}(2)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}}\frac{1}{2R_\star^2}\left[\left(\frac{\gamma^2}{2}\langle q^4\rangle_{\mathrm{c}} + 2\gamma\langle xq^2 J_1\rangle_{\mathrm{c}} + 2\langle x^2 I_2\rangle_{\mathrm{c}}\right)H_1(\nu) - \left(\frac{1}{2}\gamma^2\langle x^2q^2\rangle_{\mathrm{c}} + \frac{\gamma}{3}\langle x^3 J_1\rangle_{\mathrm{c}}\right)H_3(\nu) + \frac{\gamma^2}{24}\langle x^4\rangle_{\mathrm{c}}H_5(\nu)\right.$$
$$- \left(4\gamma\langle xq^2\rangle\langle q^2 J_1\rangle + \langle q^2 J_1\rangle\langle x^2 J_1\rangle + 4\langle xq^2\rangle\langle x I_2\rangle\right)H_1(\nu) - \left(\frac{\gamma^2}{6}\langle xq^2\rangle\langle x^3\rangle + \frac{\gamma}{6}\langle x^3\rangle\langle x^2 J_1\rangle\right)H_5(\nu) + \frac{\gamma^2}{72}\langle x^3\rangle^2 H_7(\nu)$$
$$\left. + \left(\gamma^2\langle xq^2\rangle^2 + \frac{\gamma}{3}\langle x^3\rangle\langle q^2 J_1\rangle + \gamma\langle xq^2\rangle\langle x^2 J_1\rangle + \frac{1}{4}\langle x^2 J_1\rangle^2 + \frac{2}{3}\langle x^3\rangle\langle x I_2\rangle\right)H_3(\nu)\right] \tag{A17}$$

can be simplified up to NLO to

$$\chi_{2D}^{\mathrm{iso}(0+1)}(\nu) = \frac{\sigma_1^2}{2\sigma^2}\frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}}\left[H_1(\nu) + \frac{1}{3!}H_4(\nu)\langle x^3\rangle + H_2(\nu)\langle xq^2\rangle - \gamma^{-1}\left\langle q^2 J_1\right\rangle + \mathcal{O}(\sigma^2)\right], \tag{A18}$$

which is to be compared with expression (44) for the Euler characteristic on 2D slices. The NNLO term adds

$$\chi_{2D}^{\mathrm{iso}(2)}(\nu) = \frac{\sigma_1^2}{2\sigma^2}\frac{\mathrm{e}^{-\nu^2/2}}{(2\pi)^{3/2}}\left[H_1(\nu)\left(-\frac{1}{2}\langle q^4\rangle_c - \gamma^{-1}\langle xq^2 J_1\rangle_c + \gamma^{-1}\left\langle xq^2\right\rangle\left\langle q^2 J_1\right\rangle\right)\right.$$
$$\left. + H_3(\nu)\left(\frac{1}{2}\langle x^2q^2\rangle_c - \frac{1}{6}\gamma^{-1}\langle x^3\rangle\left\langle q^2 J_1\right\rangle\right) + H_5(\nu)\left(\frac{1}{24}\langle x^4\rangle_c + \frac{1}{6}\langle xq^2\rangle\langle x^3\rangle\right) + H_7(\nu)\frac{1}{72}\langle x^3\rangle^2\right]. \tag{A19}$$

APPENDIX B: ($f_{\rm nl}, g_{\rm nl}$) NG FIELD TOY MODELS

B1 $f_{\rm nl}$ toy model for the field

Let us briefly describe the toy model used in the main text to validate our predictions for extrema counts and Minkowski functionals in 2D and 3D. If X is a Gaussian (possibly anisotropic) field with zero mean and variance $\langle X^2\rangle = \sigma_0^2$, let us define the following NG field

$$X^{\rm NG} = X + \frac{f_{\rm nl}}{\sigma_0}\left(X^2 - \sigma_0^2\right) + \frac{g_{\rm nl}}{\sigma_0^2}\left(X^3 - 3\sigma_0^2 X\right) . \tag{B1}$$

For the purpose of this paper, let us define X as a Kaiser transform of a Gaussian isotropic field, $X^{\rm G}$, which can be written in Fourier space, following equation (11), as

$$\hat{X}_k = (1 + f\mu^2)\hat{X}^{\rm G}_k, \quad \text{with} \quad \mu = k_\parallel/k. \tag{B2}$$

Then any Minkowski functional or extrema count of this NG field can simply be written in terms only of f, $f_{\rm nl}$, $g_{\rm nl}$ and σ's (or n for a scale-invariant power spectrum). Indeed, one can easily show that the spectral parameters of the NG field are the following in terms of the spectral parameters of X (in 2D): $\tilde{\sigma}^2 = (1+2f_{\rm nl}^2)\sigma_0^2$, $\tilde{\sigma}^2_{1\perp} = (1+4f_{\rm nl}^2)\sigma^2_{1\perp}$, $\tilde{\sigma}^2_{1\parallel} = (1+4f_{\rm nl}^2)\sigma^2_{1\parallel}$, $\tilde{\sigma}^2_{2\perp} = (1+4f_{\rm nl}^2+12f_{\rm nl}^2\gamma_\perp^2)\sigma^2_{2\perp}$, $\tilde{\sigma}^2_{2\parallel} = (1+4f_{\rm nl}^2+12f_{\rm nl}^2\gamma_\parallel^2)\sigma^2_{2\parallel}$ and $\tilde{\sigma}^2_Q = (1+4f_{\rm nl}^2+4f_{\rm nl}^2\gamma_\perp\gamma_\parallel/\gamma_2)\sigma^2_Q$, where in turn the spectral parameters of X can be expressed as functions of the spectral parameters of $X^{\rm G}$: $\sigma_0^2 = (1+f+3f^2/8)\sigma^2$, $\sigma^2_{1\parallel} = 1/16(8+12f+5f^2)\sigma_1^2$, $\sigma^2_{1\perp} = 1/16(8+4f+f^2)\sigma_1^2$, $\sigma^2_{2\parallel} = 1/128(48+80f+35f^2)\sigma_2^2$, $\sigma^2_{2\perp} = 1/128(48+16f+3f^2)\sigma_2^2$ and $\sigma^2_Q = 1/128(16+16f+5f^2)\sigma_2^2$. In 2D, the three-point cumulants of the NG field can also be computed: $\langle x^3\rangle = f_{\rm nl}(6+8f_{\rm nl}^2)/(1+2f_{\rm nl}^2)^{3/2}$, $\langle x q^2_{||,\perp}\rangle = 4f_{\rm nl}\sqrt{1+2f_{\rm nl}^2}/(1+4f_{\rm nl}^2)$ and $\langle J_{1\perp}q^2_{||}\rangle/\gamma_\perp = -4f_{\rm nl}\sqrt{1+2f_{\rm nl}^2}/(1+4f_{\rm nl}^2)$.

In 3D, similar relations hold: $\tilde{\sigma}^2 = (1+2f_{\rm nl}^2)\sigma_0^2$, $\tilde{\sigma}^2_{1\perp} = (1+4f_{\rm nl}^2)\sigma^2_{1\perp}$, $\tilde{\sigma}^2_{1\parallel} = (1+4f_{\rm nl}^2)\sigma^2_{1\parallel}$, $\tilde{\sigma}^2_{2\perp} = (1+4f_{\rm nl}^2+8f_{\rm nl}^2\gamma_\perp^2)\sigma^2_{2\perp}$, $\tilde{\sigma}^2_{2\parallel} = (1+4f_{\rm nl}^2+12f_{\rm nl}^2\gamma_\parallel^2)\sigma^2_{2\parallel}$ and $\tilde{\sigma}^2_Q = (1+4f_{\rm nl}^2+4f_{\rm nl}^2\gamma_\perp\gamma_\parallel/\gamma_2)\sigma^2_Q$, where again it can be related to the spectral parameters of the isotropic Gaussian field X: $\sigma_0^2 = (1+2f/3+f^2/5)\sigma^2$, $\sigma^2_{1\parallel} = (1/3+2f/5+f^2/7)\sigma_1^2$, $\sigma^2_{1\perp} = (2/3+4f/15+2f^2/35))\sigma_1^2$, $\sigma^2_{2\parallel} = (1/5+2f/7+f^2/9)\sigma_2^2$, $\sigma^2_{2\perp} = (8/15+16f/105+8f^2/315)\sigma_2^2$ and $\sigma^2_Q = (2/15+4f/35+2f^2/63)\sigma_2^2$, the three-point cumulants of interest being: $\langle x^3\rangle = f_{\rm nl}(6+8f_{\rm nl}^2)/(1+2f_{\rm nl}^2)^{3/2}$, $\langle x q^2_{||,\perp}\rangle = 4f_{\rm nl}\sqrt{1+2f_{\rm nl}^2}/(1+4f_{\rm nl}^2)$, $\langle J_{1\perp}q^2_\perp\rangle/\gamma_\perp = -2f_{\rm nl}\sqrt{1+2f_{\rm nl}^2}/(1+4f_{\rm nl}^2)$ and $\langle J_{1\perp}q^2_{||}\rangle/\gamma_\perp = -4f_{\rm nl}\sqrt{1+2f_{\rm nl}^2}/(1+4f_{\rm nl}^2)$.

Figs B1 and B2 illustrate the interplay between anisotropy and non-Gaussianity that this paper investigates on this example of the '$f_{\rm nl}$' toy model. It presents the 2D/3D Euler characteristic of a mildly NG anisotropic field for different values of the quadratic parameter $f_{\rm nl}$ and anisotropy, f. Both predictions derived in Sections 4.2.1 and 4.2.2, and measurements from Monte Carlo realizations of the corresponding field are shown. The agreement between theory and measurements is very good. As expected, in 2D the effect of anisotropy is larger than in 3D.

Fig. B3 shows how $g_{\rm nl}$ acts in turn. For this purpose, the prediction at NNLO ($n = 4$ in the GC expansion) is computed from equation (40). Finally, Fig. B4 displays predictions and measurements for differential extrema counts in these kinds of $f_{\rm nl}$ fields. As mentioned in the main text, at even orders in the GC expansion (in particular at Gaussian order), the 2D saddle count is even, peak and void counts are symmetric by reflection across the y-axis meaning $\partial_\nu n_{\rm peaks}(\nu) = \partial_\nu n_{\rm voids}(-\nu)$ and pancake and filament counts are also symmetric by reflection across the y-axis. On the other hand, at odd orders in the GC expansion (in particular at first NG order), the 2D saddle count is odd in ν, peak and void counts are symmetric w.r.t. reflection relative to the origin, meaning $\partial_\nu n_{\rm peaks}(\nu) = -\partial_\nu n_{\rm voids}(-\nu)$ and pancake and filament counts are also symmetric by reflection relative to the origin. Fig. B4 thus displays in left-hand panels the even part of the data which should be dominated by the Gaussian term, namely $(\partial_{\ nu} n_{\rm sad}(\nu) + \partial_\nu n_{\rm sad}(-\nu))/2$ for 2D saddle points, $(\partial_\nu n_{\rm peaks}(\nu) + \partial_\nu n_{\rm voids}(-\nu))/2$ for peaks and $(\partial_\nu n_{\rm voids}(\nu) + \partial_\nu n_{\rm peaks}(-\nu))/2$ for voids (pancakes and filaments accordingly); and in the right-hand panels, the even part of the measurement, namely $(\partial_\nu n_{\rm sad}(\nu) - \partial_\nu n_{\rm sad}(-\nu))/2$ for 2D saddle points, $(\partial_\nu n_{\rm peaks}(\nu) - \partial_\nu n_{\rm voids}(-\nu))/2$ for peaks and $(\partial_\nu n_{\rm voids}(\nu) - \partial_\nu n_{\rm peaks}(-\nu))/2$ for voids. Agreement between measurement and theory to NLO is very good. The small difference between both is due to NNLO correction and to a bias in the measurement of pancake counts which is yet to be fixed.

B2 $f_{\rm nl}$ model for the potential

In astrophysics, $f_{\rm nl}$ often refers to the potential from which the (density) field derives. Let us thus study in this section the effect of such $f_{\rm nl}$ non-Gaussianities in the potential on the genus of the density field. If Φ is a Gaussian (possibly anisotropic) potential field with zero mean and variance $\langle\Phi^2\rangle = \sigma_0^2$, let us define the following NG field

$$\Phi^{\rm NG} = \Phi + \frac{f_{\rm nl}}{\sigma_0}\left(\Phi^2 - \sigma_0^2\right) . \tag{B3}$$

Then the resulting NG density field reads

$$\delta^{\rm NG} = \Delta\Phi^{\rm NG} = \Delta\Phi + 2\frac{f_{\rm nl}}{\sigma_0}\left(\Phi\Delta\Phi + \nabla\Phi\cdot\nabla\Phi\right) . \tag{B4}$$

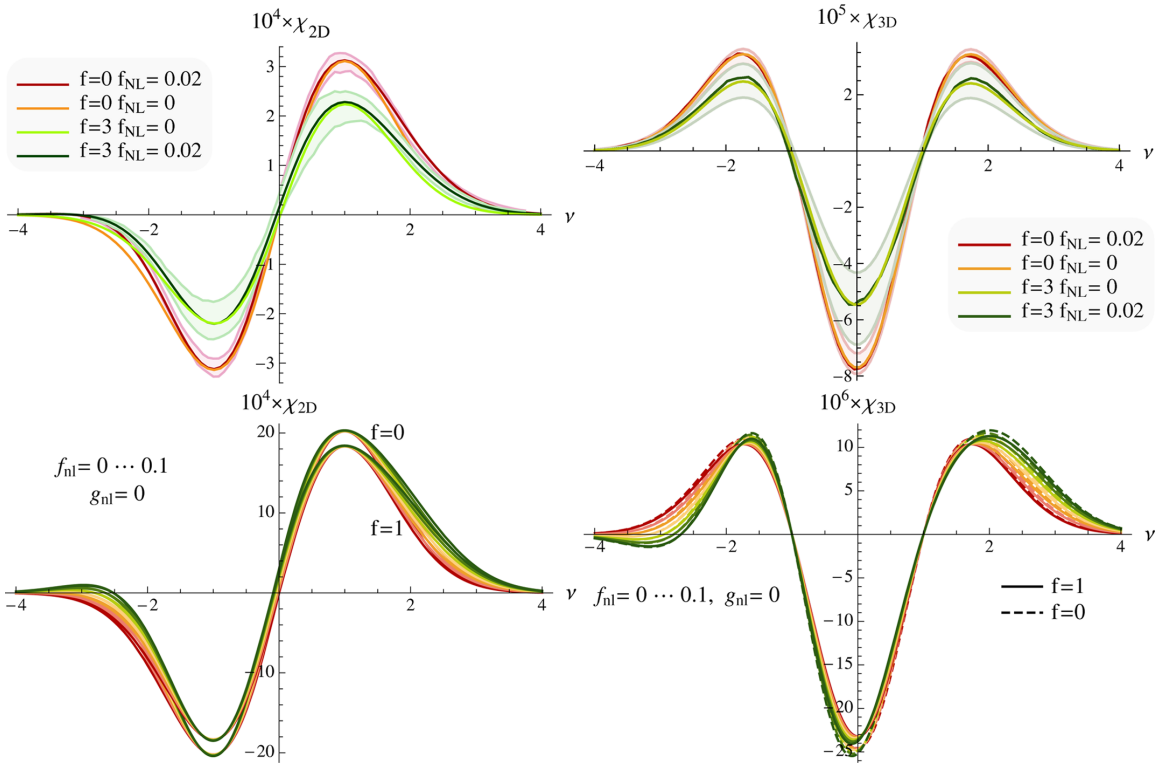


Figure B1. 2D (left) and 3D (right) Euler characteristic of a mildly NG anisotropic field in a toy model. Top left: comparison with measurements for different values of the quadratic parameter f_{nl} and anisotropy f, as labelled. The shaded area corresponds to the one-sigma measurement error on realizations of the NG fields. The overall amplitude has been rescaled by 10^4. Note both the change in amplitude with f and the distortion in the tails of the distribution with f_{nl} which are well fitted by the model. Bottom-left panel: theoretical evolution of the 2D Euler characteristic as a function of f_{nl} as labelled for $f = 0$ and 1. For this level of anisotropy, the difference in amplitude is less significant. Top-right and bottom-right panels: same as the corresponding left-hand panels in 3D.

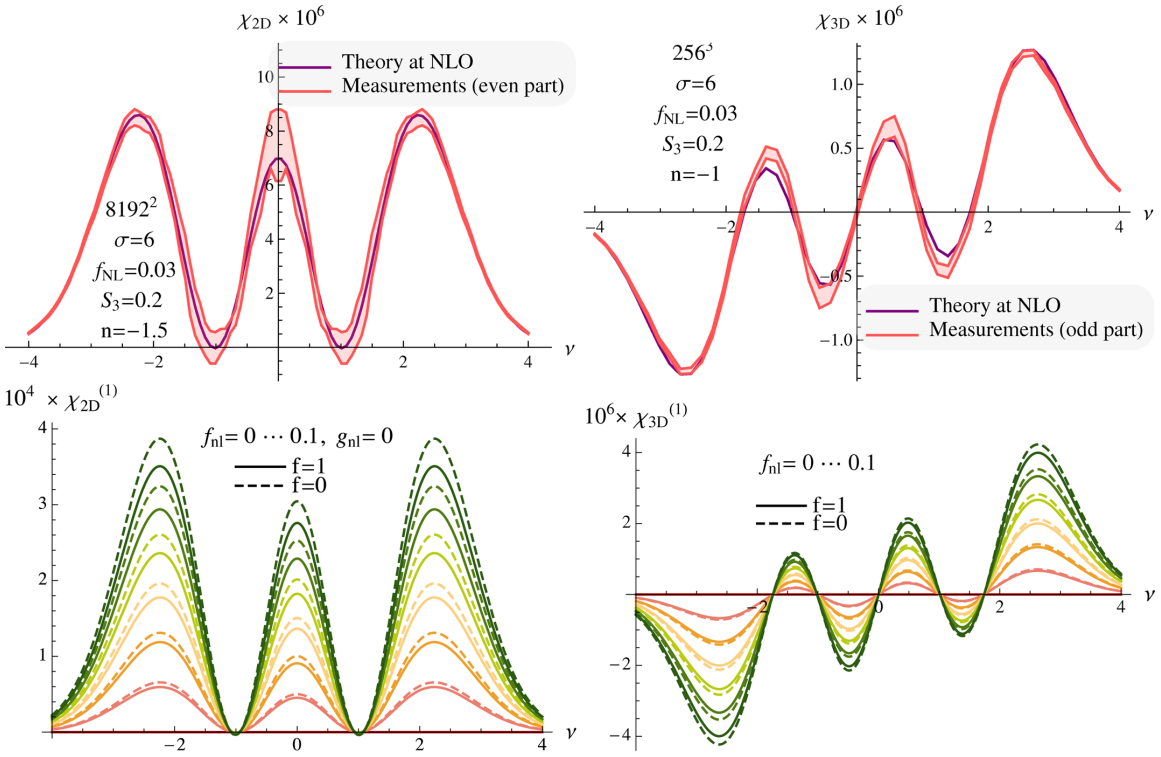


Figure B2. NG part of the Euler characteristic of a mildly NG anisotropic field in 2D (left-hand panels) and 3D (right-hand panels). The top panels compare the theoretical prediction to the measurements of realizations of the NG fields with parameters as labelled, while the bottom panels illustrate the variation of the theory with the amplitude of f_{nl}. The shaded area corresponds to the one-sigma measurement error. The match to the theory is very good throughout. The effect of redshift distortion is clearly weaker in 3D, as expected.

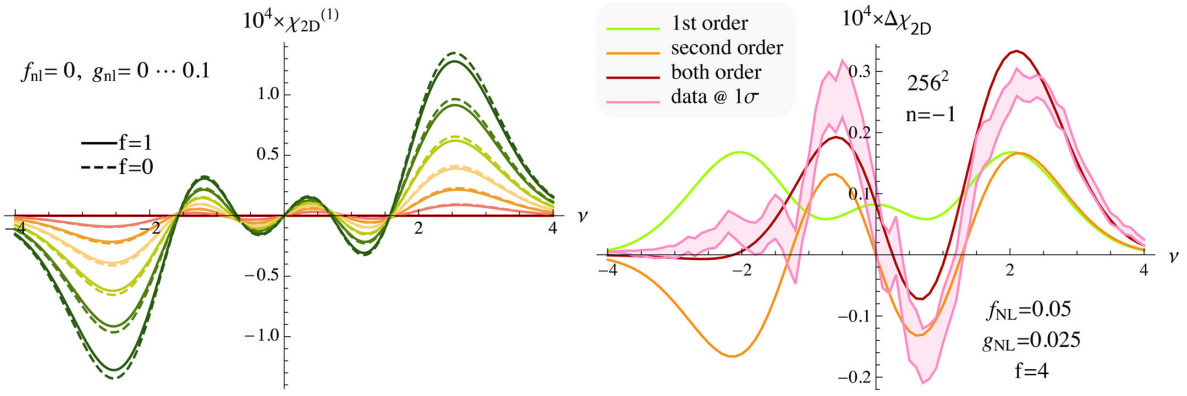


Figure B3. Left: variation with the amplitude of g_{nl} of the predicted Euler characteristic at NLO of a mildly 2D NG anisotropic field. Right: match between a given realization with f_{nl} and g_{nl} as labelled, and the corresponding model decomposed into first-order ($n = 3$) and second-order ($n = 4$) corrections.

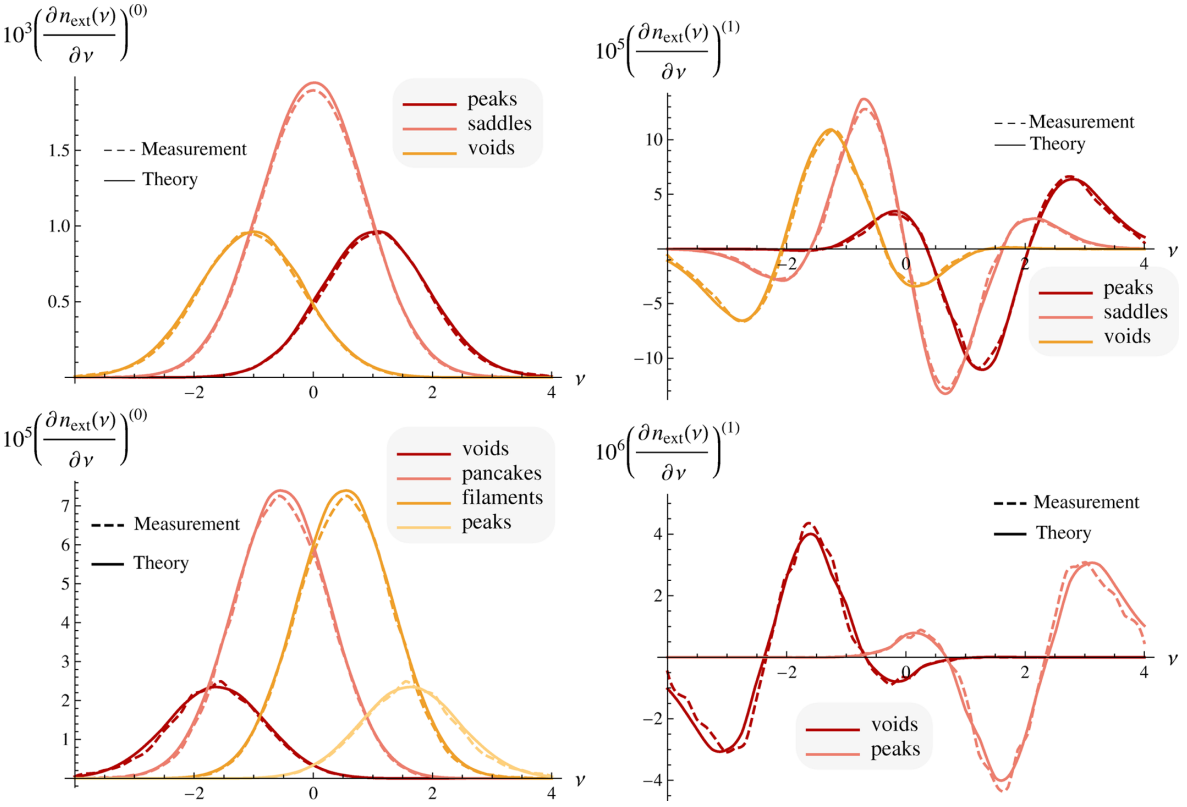


Figure B4. Top-left panel: the distribution of 2D extrema (voids, saddles and peaks as labelled) as a function of threshold ν in redshift space for $f = 1$ as predicted at Gaussian order (solid) and measured ('even part' of the data only, see the main text for more details). Top-right panel: the corresponding NG correction ($f_{nl} = 0.05$) estimated from 20 realizations of the corresponding scale-invariant ($n = -1$) field of dimension 4096^2 smoothed over 4 pixels as predicted at first NG order (solid) and measured ('odd part' of the data only). Bottom panels: same as the top panels for the distribution of 3D extrema (voids, wall saddles, tube saddle and peaks as labelled) for NG scale-invariant ($n = -1$, 256^3 smoothed over 4 pixels) anisotropic field ($f_{nl} = 0.07, f = 1$).

For the purpose of this paper, let us define again the final density field as a Kaiser transform of this field (cf. equation 11) so that in Fourier space

$$\hat{\delta}_k = (1 + f\mu^2)\hat{\delta}_k^{\rm NG}, \quad \text{with} \quad \mu = k_\parallel/k. \tag{B5}$$

Then any Minkowski functional or extrema count of this NG field can be written in terms only of f, f_{nl} and σ's (or n for a scale-invariant power spectrum). Fig. B5 (top panels) shows the prediction for the 2D Euler characteristic for different values of f_{nl} between 0 and 1 and for $f = 0$ (isotropic case) and $f = 1$ (Kaiser). It also compares (bottom panels) the prediction with some measurements and shows great agreement up to first order (the difference we observe must come from higher order corrections as the parity in Hermite polynomials is different).

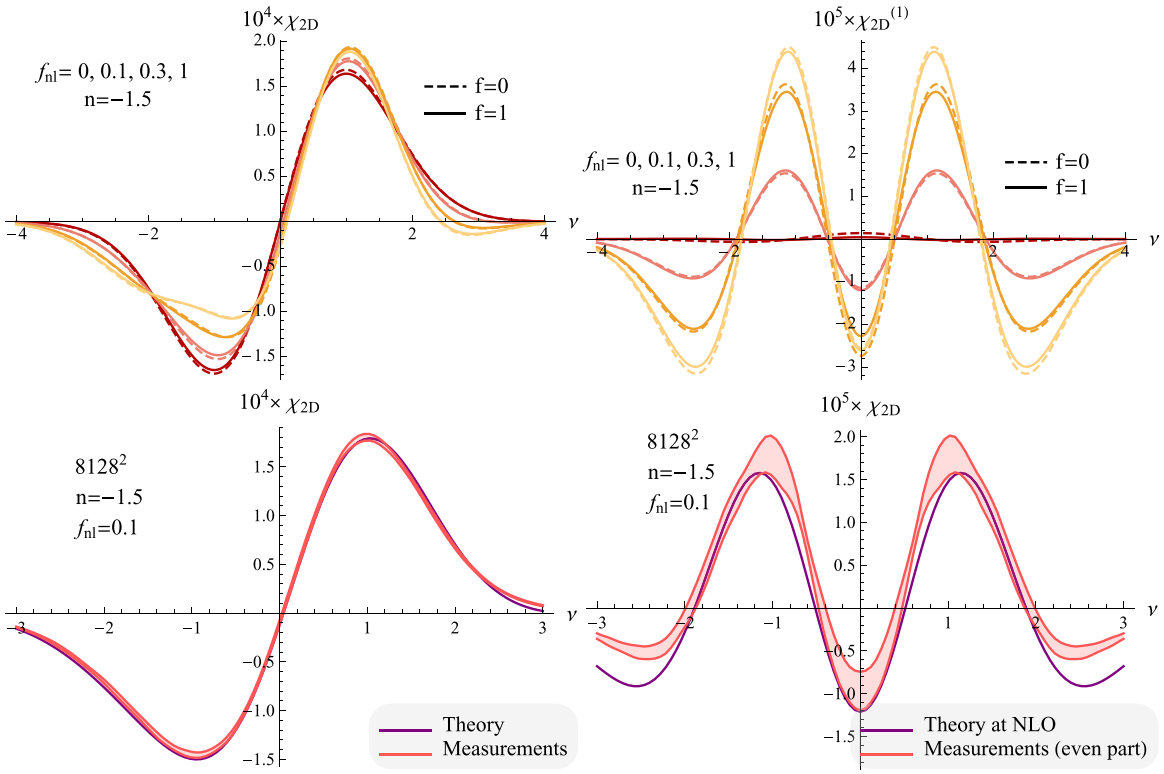


Figure B5. Top left: 2D Euler characteristic of the density field for $f_{\rm nl}$ between 0 (red) and 1 (yellow) in the potential field and $f = 0$ (dashed line) or 1 (solid line). Top right: same as the top-left panel for the first NG correction. Bottom left: 2D Euler characteristic of the density field for $f_{\rm nl} = 0.1$ in the potential field. Measurements are displayed in pink with a shaded area representing the 1σ dispersion and prediction in violet. Bottom right: same as the left-hand panel for the first NG correction.

APPENDIX C: 3D EULER CHARACTERISTIC AT ALL ORDERS IN NON-GAUSSIANITY

To compute the 3D Euler characteristics, one needs to take into account the variable Υ. Given that the GC polynomial expansion for the NG PDF is obtained by using polynomials orthogonal with respect to the kernel provided by the JPDF in the Gaussian limit, the coupling between Υ and $Q^2, J_{2\perp}$ variables in the Gaussian limit, introduces a set of polynomials F_{lrs} such that

$$\frac{1}{\pi}\int \mathrm{d}J_{2\perp}\,\mathrm{d}Q^2\,\frac{\mathrm{d}\Upsilon}{\sqrt{Q^4 J_{2\perp}-\Upsilon^2}}\mathrm{e}^{-J_{2\perp}-Q^2}F_{lrs}\left(J_{2\perp},Q^2,\Upsilon\right)F_{l'r's'}\left(J_{2\perp},Q^2,\Upsilon\right)=\delta_{ll'}\delta_{rr'}\delta_{ss'}. \tag{C1}$$

Here l is the power of $J_{2\perp}$, r the power of Q^2 and s the power of Υ. In particular, $F_{lr0}(J_{2\perp}, Q^2, \Upsilon) = L_l(J_{2\perp})L_r(Q^2)$ and $F_{001}(J_{2\perp}, Q^2, \Upsilon) = H_1(\Upsilon)$.

The full expansion of the NG JPDF $P = P(x, q_\perp^2, x_3, \zeta, J_{2\perp}, \xi, Q^2, \Upsilon)$ is a series of products of F_{lrs} and familiar Hermite and Laguerre polynomials for the rest of the variables, namely

$$P = G\left[1+\sum_{n=3}^{\infty}\sum_{\sigma_n}\frac{(-1)^{j+l+r}g_{lrs}}{i!\,j!\,k!\,l!\,m!\,r!\,p!\,s!}\left\langle x^i q_\perp^{2j}\zeta^k x_3^m\xi^p J_{2\perp}^l Q^{2r}\Upsilon^s\right\rangle_{\rm GC} H_i(x)L_j\left(q_\perp^2\right)H_k(\zeta)H_m(x_3)H_p(\xi)F_{lrs}\left(J_{2\perp},Q^2,\Upsilon\right)\right], \tag{C2}$$

where $\sigma_n = \{(i, j, k, l, m, p, r, s) \in \mathbb{N}^7 | i + 2j + k + 2l + m + 2r + p + 3s = n\}$ and G is given by equation (18). The terms of the expansion (C2) are sorted in the order of power in field variables n. We immediately see that at $n = 3$ order, the only contribution from Υ comes from the term corresponding to $i = j = l = m = p = r = 0, s = 1$.

The GC coefficients are defined by

$$\left\langle x^i q_\perp^{2j}\zeta^k x_3^m\xi^p J_{2\perp}^l Q^{2r}\Upsilon^s\right\rangle_{\rm GC} = \frac{(-1)^{j+l+r}}{g_{lrs}}\,j!\,l!\,r!\,s!\left\langle H_i(x)L_j\left(q_\perp^2\right)H_k(\zeta)H_m(x_3)H_p(\xi)F_{lrs}\left(J_{2\perp},Q^2,\Upsilon\right)\right\rangle, \tag{C3}$$

where g_{lrs} is such that the highest power term in $F_{lrs}(J_{2\perp}, Q^2, \Upsilon)$ is $(-1)^{l+r}g_{lrs}J_{2\perp}^l Q^{2r}\Upsilon^s/(l!r!s!)$. Amongst them, we have in particular $g_{lr0} = 1$ and $g_{001} = 1$. From equation (31), it is clear that the 3D Euler characteristic is the sum of a term where Upsilon can be marginalized over (so that the formalism introduced in Section 3 can be used) and a term which integrates Υ under the condition of zero gradient. The

latter part is simply computed because $\Upsilon = F_{001}(J_{2\perp}, Q^2, \Upsilon)$ is orthogonal to all other F_{lrs} polynomials in the GC expansion of the JPDF. After some algebra, it is found that the full 3D Euler characteristic can be eventually written

$$\chi_{3\mathrm{D}}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{8\pi^2}\frac{\sigma_{1\parallel}\sigma_{1\perp}^2}{\sigma^3}H_2(\nu) + \sum_{n=3}^{\infty}\chi_{3\mathrm{D}}^{(n)}, \tag{C4}$$

with $\chi_{3\mathrm{D}}^{(n)}$ given by

$$\chi_{3\mathrm{D}}^{(n)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{8\pi^2}\frac{\sigma_{2\perp}^2\sigma_{2\parallel}}{\sigma_{1\perp}^2\sigma_{1\parallel}}\left[\sum_{\sigma_n}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_{i+2}(\nu)\gamma_\parallel\gamma_\perp^2\left\langle x^i q_\perp^{2j}x_3^{2m}\right\rangle_{\mathrm{GC}} - \sum_{\sigma_{n-1}}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_{i+1}(\nu)\left\langle x^i q_\perp^{2j}x_3^{2m}\left(\gamma_\perp^2 x_{33} + 2\gamma_\perp\gamma_\parallel J_{1\perp}\right)\right\rangle_{\mathrm{GC}}\right.$$
$$\left. + \sum_{\sigma_{n-2}}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_i(\nu)\left\langle x^i q_\perp^{2j}x_3^{2m}\left(2\gamma_\perp(J_{1\perp}x_{33} - \gamma_2 Q^2) + \gamma_\parallel(J_{1\perp}^2 - J_{2\perp})\right)\right\rangle_{\mathrm{GC}} - \sum_{\sigma_{n-3}}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_{i-1}(\nu)\left\langle x^i q_\perp^{2j}x_3^{2m}\,(4I_3)\right\rangle_{\mathrm{GC}}\right], \tag{C5}$$

where $I_3 = \frac{1}{4}x_{33}(J_{1\perp}^2 - J_{2\perp}) - \frac{1}{2}\gamma_2(Q^2 J_{1\perp} - \Upsilon)$. In equation (C5), the GC coefficients are written in a concise form but must be interpreted returning from the $(x, x_{33}, J_{1\perp})$ to the (x, ζ, ξ) variables and using equation (20). In the isotropic limit, the result is reduced to the following compact form

$$\chi_{3\mathrm{D}}^{(n)}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{8\pi^2}\frac{\sigma_1^3}{\sigma^3}\left[\frac{2}{3\sqrt{3}}\sum_{\sigma_n}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_{i+2}(\nu)\left\langle x^i q_\perp^{2j}x_3^{2m}\right\rangle_{\mathrm{GC}} - \frac{2}{\sqrt{3}}\gamma^{-1}\sum_{\sigma_{n-1}}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_{i+1}(\nu)\left\langle x^i q_\perp^{2j}x_3^{2m}I_1\right\rangle_{\mathrm{GC}}\right.$$
$$\left. + 2\sqrt{3}\gamma^{-2}\sum_{\sigma_{n-2}}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_i(\nu)\left\langle x^i q_\perp^{2j}x_3^{2m}I_2\right\rangle_{\mathrm{GC}} - \frac{16}{5}\sqrt{\frac{3}{5}}\gamma^{-3}\sum_{\sigma_{n-3}}\frac{(-1)^{j+m}}{2^m i!\,j!\,m!}H_{i-1}(\nu)\left\langle x^i q_\perp^{2j}x_3^{2m}I_3\right\rangle_{\mathrm{GC}}\right], \tag{C6}$$

where I_1, I_2 and I_3 are the three invariants of the Hessian matrix x_{ij}, namely $I_1 = J_1 = \lambda_1 + \lambda_2 + \lambda_3 = \mathrm{Tr}\,x_{ij}$, $I_2 = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1$ and $I_3 = \lambda_1\lambda_2\lambda_3 = \det x_{ij}$. Note that equations (C5) and (C6) contain some $H_{-1}(\nu) = \sqrt{\pi/2}\,\mathrm{Erfc}(\nu/\sqrt{2})$ which are shown to contribute as a boundary term and can thus be factorized out of the terms $\chi_{3\mathrm{D}}^{(n)}$ as mentioned in equation (33). Moreover, the well-known topological relation between the Euler characteristics and the genus allows us to compare the above quantity with the curvature integrated along isocontours (as used in Matsubara (1996)). In an anisotropic space, this surface integral is found to be

$$\chi_{3\mathrm{D}}^{\mathrm{s}}(\nu) = \frac{\mathrm{e}^{-\nu^2/2}}{8\pi^2}\frac{\sigma_{1\|}\sigma_{1\perp}^2}{\sigma^3}\left[H_2(\nu) + \frac{1}{\gamma_\perp^2}\sum_{n=3}^{\infty}\sum_{\sigma_{n-2}}\frac{(-1)^{j+m}}{i!\,j!m!(2m-1)2^m}H_i(\nu)\left(\left\langle x^i q_\perp^{2j}J_{2\perp}\,x_3^{2m}\right\rangle_{\mathrm{GC}} - (1-\gamma_\perp^2)\left\langle x^i q_\perp^{2j}\zeta^2 x_3^{2m}\right\rangle_{\mathrm{GC}}\right)\right.$$
$$+ 2\frac{\sqrt{1-\gamma_\perp^2}}{\gamma_\perp}\sum_{n=3}^{\infty}\sum_{\sigma_{n-1}}\frac{(-1)^{j+m}}{i!\,j!m!(2m-1)2^m}\left\langle x^i q_\perp^{2j}\zeta\,x_3^{2m}\right\rangle_{\mathrm{GC}}H_i(\nu)H_1(\nu)$$
$$\left. - \sum_{n=3}^{\infty}\sum_{\sigma_n}\frac{(-1)^{j+m}}{i!\,j!m!(2m-1)2^m}\left\langle x^i q_\perp^{2j}x_3^{2m}\right\rangle_{\mathrm{GC}}H_i(\nu)H_2(\nu)\right], \tag{C7}$$

where we use the convention $(-2)!/(-1)! = -1/2$. Equations (C4) and (C7) were checked to be equivalent up to at least $n = 4$ using some detailed relations between the cumulants (which once again can be established via integrations by parts) assuming that everything vanishes at infinity [in particular $\chi_{3\mathrm{D}}(-\infty) = 0$]. Note that in this expression ξ, Q^2 and Υ do not appear anymore.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed September 16, 2015 (MN LATEX style file v2.2)

The realm of dark halos: peak-peak exclusion zone

Tobias Baldauf[1], Sandrine Codis[2], Vincent Desjacques[3] and Christophe Pichon[2,4]

[1] *Institute for Advanced Study, School of Natural Sciences, Einstein Drive, Princeton, NJ 08540, USA*
[2] *CNRS, UMR7095 & UPMC Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, 75014, Paris, France*
[3] *Département de Physique Théorique. Université de Genève, 24, quai Ernest Ansermet. 1211, Genève, Switzerland*
[4] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, United Kingdom*

September 16, 2015

ABSTRACT

The small-scale exclusion in the two-halo term and the convergence of perturbative bias expansions is presented. Peaks of a 1D cosmological random field are used as a proxy for a catalogue of biased tracers. Starting from the two-point correlation function of peaks of a given height, series approximations are derived that are valid inside the exclusion zone and emphasize deviations from Poisson noise in the power spectrum at low frequencies. The convergence of the perturbative bias expansion is discussed. Finally, we go beyond Gaussian statistics for the initial conditions and investigate the subsequent evolution of peaks through their Zel'dovich ballistic displacement. Even though these findings apply to the clustering of one-dimensional tracers, they provide useful insights into halo exclusion and its impact on the two-halo term.

Key words:

1 INTRODUCTION

Dark matter haloes and the galaxies within them are distinct and extended objects. By definition, they can not overlap and their centers have to be separated by at least the sum of their virial radii. This exclusion effect is even more important in the initial conditions or Lagrangian space, before the objects collapsed and fell towards each other. As noted in Mo & White (1996); Sheth & Lemson (1999), the vanishing probability to find two centers closer than the exclusion radius corresponds to the correlation function being -1 for small separations. While this effect is localized at small separations in the correlation function, it can alter the power spectrum at small wavenumbers and results in a modification of Poisson stochasticity (Smith et al. 2007; Baldauf et al. 2013) consistent with the sub-Poissonian noise measured in the clustering of simulated dark matter halos (Casas-Miranda et al. 2002; Seljak et al. 2009; Hamaus et al. 2010; Manera & Gaztañaga 2011). Exclusion effects strongly suppress the non-physical k^0 tail of the one-halo term (Smith et al. 2011).

Besides the exclusion effects, there are distinct, non-linear bias effects just outside the exclusion region, which, due to their localization, also contribute to the power spectrum on large scales and for which the bias expansion converges very slowly.

The exclusion region and the non-linear bias bump beyond are important for precision models of the halo-halo correlation function in the transition region between the one- and two-halo terms in the halo model. Thus, a better understanding of these regions will likely improve the modelling of the matter power spectrum or correlation function in this regime, which is very important for weak lensing or galaxy-galaxy lensing studies, since this signal is large and not yet dominated by the fully non-perturbative one-halo term.

Dark matter haloes are seeded by over-dense regions in Lagrangian space (so called proto-haloes) that subsequently collapse to form the virialized late-time Eulerian haloes. Various assumptions can be made to describe the relation between the proto-haloes and the underlying Gaussian density field. Our perfect knowledge of the N-point statistics of the Gaussian field allows us to calculate all possible statistics of transformations of the Gaussian field. In this paper we will consider the peak model, in which proto-haloes are associated with the maxima of the smoothed underlying field.

To simplify the calculations and understanding, but without loosing much of the phenomenology, we will consider volume exclusion effects associated with peaks in one spatial dimension, following Lumsden et al. (1989); Coles (1989).

The paper is organized as follows. In Sec. 2, the mathematical formalism that defines peak-peak correlations functions

together with their numerical implementation are described. Sec. 3 presents the result on the small-scale exclusion zone of peaks obtained by numerical integration. The analytical large-scale expansion of the two-point correlation function of peaks is then discussed in Sec. 4. Sec. 5 incorporates the effect of the Zel'dovich displacement of the peaks. Finally, Sec. 6 wraps up.

2 FORMALISM AND NUMERICAL IMPLEMENTATION

The formalism of cosmological density peaks, which builds on the Kac-Rice formula Kac (1943); Rice (1945) was laid down in Bardeen et al. (1986). Following Pogosyan et al. (2009), for a given field ρ, we define the moments

$$\sigma_0{}^2 = \langle \rho^2 \rangle, \qquad \sigma_1{}^2 = \langle (\nabla\rho)^2 \rangle, \qquad \sigma_2{}^2 = \langle (\Delta\rho)^2 \rangle. \tag{1}$$

Combining these moments, we can build two characteristic lengths $R_0 = \sigma_0/\sigma_1$ and $R_\star = \sigma_1/\sigma_2$, as well as the spectral parameter

$$\gamma = \frac{\sigma_1{}^2}{\sigma_0\sigma_2}. \tag{2}$$

We choose to normalise the field and its derivatives to have unit variances:

$$x = \frac{1}{\sigma_0}\rho, \qquad x_i = \frac{1}{\sigma_1}\nabla_i\rho, \qquad x_{ij} = \frac{1}{\sigma_2}\nabla_i\nabla_j\rho. \tag{3}$$

In general, while $\mathcal{P}(\mathbf{X})$ designates the one-point probability density (PDF), $\mathcal{P}(\mathbf{X},\mathbf{Y})$ will denote the joint PDF for the normalized field and its derivatives, $\mathbf{X} = \{x, x_{ij}, x_i\}$ and $\mathbf{Y} = \{y, y_{ij}, y_i\}$, at two prescribed comoving locations ($\mathbf{r}_x$ and $\mathbf{r}_y$ separated by a distance $r = |\mathbf{r}_x - \mathbf{r}_y|$). In the particular case of Gaussian initial conditions, this joint PDF is the multivariate Normal

$$\mathcal{N}(\mathbf{X},\mathbf{Y}) = \frac{\exp\left[-\frac{1}{2}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}^{\mathrm{T}}\cdot\mathbf{C}^{-1}\cdot\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}\right]}{\det|\mathbf{C}|^{1/2}\,(2\pi)^{(\mathrm{D}+1)(\mathrm{D}+2)/2}}, \tag{4}$$

where $\mathbf{C}_0 \equiv \langle \mathbf{X}\cdot\mathbf{X}^{\mathrm{T}}\rangle$ and $\mathbf{C}_\gamma \equiv \langle \mathbf{X}\cdot\mathbf{Y}^{\mathrm{T}}\rangle$ are the diagonal and off-diagonal components of the covariance matrix

$$\mathbf{C} = \begin{pmatrix}\mathbf{C}_0 & \mathbf{C}_\gamma\\ \mathbf{C}_\gamma^{\mathrm{T}} & \mathbf{C}_0\end{pmatrix}. \tag{5}$$

All these quantities depend on the separation vector $\mathbf{r}$ only because of homogeneity. Isotropy further implies that they depend on the modulus $r = |\mathbf{r}|$ solely. Eq.(4) is sufficient to compute the expectation of any quantity involving the fields and its derivatives up to second order. In particular, the two-point correlation $\xi_{\mathrm{crit}}(r,\nu)$ of (signed) critical points at threshold ν separated by r is given by

$$1 + \xi_{\mathrm{crit}}(r,\nu) = \frac{\left\langle n_{\mathrm{crit}}(\mathbf{X}) n_{\mathrm{crit}}(\mathbf{Y})\right\rangle}{\left\langle n_{\mathrm{crit}}(\mathbf{X})\right\rangle^2}, \tag{6}$$

where the Klimontovich or "localized" density for a signed critical point reads

$$n_{\mathrm{crit}}(\mathbf{X}) = \left(\frac{\sigma_2}{\sigma_1}\right)^{\mathrm{D}} \det(x_{ij})\delta_{\mathrm{D}}(x_i)\delta_{\mathrm{D}}(x-\nu). \tag{7}$$

This density is formally zero unless the condition for a critical point is satisfied. The multiplicative factor of $(\sigma_1/\sigma_2)^{\mathrm{D}}$, which has dimension of length$^{-\mathrm{D}}$, ensures that the ensemble average

$$\left\langle n_{\mathrm{crit}}(\mathbf{X})\right\rangle = \int \mathrm{d}\mathbf{X}\det(x_{ij})\delta_{\mathrm{D}}(x_i)\delta_{\mathrm{D}}(x-\nu)\mathcal{P}(\mathbf{X}) \equiv \bar{n}_{\mathrm{crit}}(\nu), \tag{8}$$

which appears in the denominator of Eq.(6), equals the average number density of critical points at threshold ν. The ensemble average

$$\left\langle n_{\mathrm{crit}}(\mathbf{X}) n_{\mathrm{crit}}(\mathbf{Y})\right\rangle = \int \mathrm{d}\mathbf{X}\int \mathrm{d}\mathbf{Y}\,\mathcal{P}(\mathbf{X},\mathbf{Y})\det(x_{ij})\delta_{\mathrm{D}}(x_i)\delta_{\mathrm{D}}(x-\nu)\det(y_{ij})\delta_{\mathrm{D}}(y_i)\delta_{\mathrm{D}}(y-\nu) \tag{9}$$

is the cross-correlation. Since the integrant simply is a polynomial function of the variables, this integral can be fully carried out analytically. For peaks, an additional constraint on the sign of the second derivatives is required. As a consequence, the peak two-point correlation becomes

$$1 + \xi_{\mathrm{pk}}(r,\nu) = \frac{\left\langle n_{\mathrm{pk}}(\mathbf{X}) n_{\mathrm{pk}}(\mathbf{Y})\right\rangle}{\left\langle n_{\mathrm{pk}}(\mathbf{X})\right\rangle^2}. \tag{10}$$

where the localized peak number density $n_{\mathrm{pk}}(\mathbf{X})$,

$$n_{\mathrm{pk}}(\mathbf{X}) = \left(\frac{\sigma_2}{\sigma_1}\right)^{\mathrm{D}} |\det(x_{ij})|\delta_{\mathrm{D}}(x_i)\Theta_{\mathrm{H}}(-\lambda_i)\delta_{\mathrm{D}}(x-\nu), \tag{11}$$

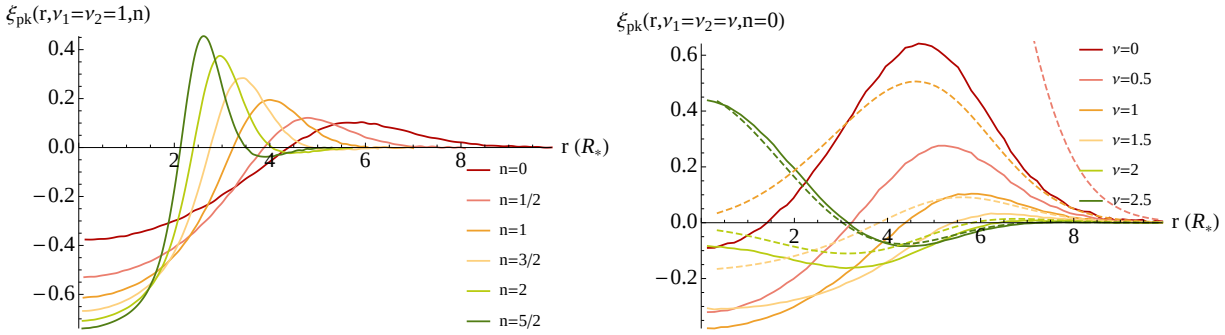


Figure 1. *Left panel*: 1D correlation function of peaks of height $\nu = 1$ as a function of $r/R_\star$ for different spectral index n evaluated by Monte-Carlo realisations of the peak constraint. *Right panel*: Same as left panel, but the spectral index is held fixed at $n = 0$ while the peak height ν is varied between 0 and 2.5 as labeled. The dashed lines indicate the analytical 1D *signed* critical point correlation function. For $\nu > 2$, both the peak and signed critical points correlations are in very good agreement. Note that, for $\nu = 0$, the signed critical point correlation functions diverges and is thus not shown on this plot.

implements the peak condition. Here $|\det(x_{ij})| = -\det(x_{ij})$ because the determinant is negative at the peaks and it is understood that, for $D > 1$, $\delta_{\rm D}(x_i)$ stands for $\prod_{i\leqslant D}\delta_{\rm D}(x_i)$, while $\Theta_{\rm H}(-y_{ii})$ means $\prod_{l\leqslant D}\Theta_{\rm H}(-\lambda_l)$ with $\{\lambda_l\}_l$ the eigenvalues of the Hessian. Because of these inequalities, the integral typically is not analytical anymore. In dimension D, we define the conditional probability that x_{ij} and y_{ij} satisfy the PDF, subject to the condition that $x_i = 0$ and $x = y = \nu$ and resort to Monte-Carlo methods in MATHEMATICA in order to evaluate numerically equation (10). Namely, we draw random numbers of dimension D(D + 1) from the conditional probability that x_{ij} and y_{ij} satisfy the PDF, subject to the condition that $x_i = 0$ and $x = y = \nu$ (using `RandomVariate`). For each draw $^{(k)}$ if $\lambda_l(x_{ij}^{(k)}) < 0$ and $\lambda_l(y_{ij}^{(k)}) < 0$ $(l \leqslant D)$ we keep the sample and evaluate $\det(x_{ij}^{(k)})\det(y_{ij}^{(k)})$ and otherwise we drop it; eventually,

$$\left\langle n_{\rm pk}(\mathbf{X})n_{\rm pk}(\mathbf{Y})\right\rangle \approx \frac{1}{N}\sum_{k\in\mathcal{S}}\left[\det(x_{ij}^{(k)})\det(y_{ij}^{(k)})\right]\times\mathcal{P}(x=y=\nu,x_i=y_i=0)\,, \tag{12}$$

where N is the total number of draws, and $\mathcal{S}$ is the subset of the indexes of draws satisfying the constraints on the eigenvalues. The same procedure can be applied to evaluate the denominator $\langle n_{\rm pk}(\mathbf{X})\rangle \equiv \bar{n}_{\rm pk}(\nu)$. Equation (10) then yields $\xi_{\rm pk}(r,\nu)$. This algorithm is embarrassingly parallel and can be easily generalized, for instance, to the computation of the correlation function $\xi_{\rm pk}(r,>\nu)$ of peaks *above* a given threshold in density and to higher dimension $D > 1$. In practice it is fairly efficient as the draw is customized to the shape of the underlying Gaussian PDF. For D= 1 considered in this work, this brute force Monte-Carlo method converges relatively quickly, typically we use one million draws for each evaluation of the correlation function in this work, this number is sufficient to reach percent precision accuracy. Obviously if correlation function above a given threshold are considered, the required number of draws is larger and increases with the value of the threshold (as the event $x > \nu$ becomes rarer).

3 SMALL SCALES: PEAK-PEAK EXCLUSION

Here and henceforth, we will assume Gaussian initial conditions such that $\mathcal{P}(\mathbf{X},\mathbf{Y})$=$\mathcal{N}(\mathbf{X},\mathbf{Y})$. In 1D, the block matrices that make up the covariance matrix are

$$\mathbf{C}_0 = \begin{pmatrix} 1 & 0 & -\gamma \\ 0 & 1 & 0 \\ -\gamma & 0 & 1 \end{pmatrix}\,,\quad \mathbf{C}_\gamma = \begin{pmatrix} \gamma_{00} & \gamma_{01} & \gamma_{02} \\ \gamma_{01} & \gamma_{11} & \gamma_{12} \\ \gamma_{02} & \gamma_{12} & \gamma_{22} \end{pmatrix}\,, \tag{13}$$

where the γ_{ij}'s represent the correlations between the field and its derivatives at two points separated by a comoving distance r, e.g. $\gamma_{22} = \langle x_{11}y_{11}\rangle$. These γ_{ij}'s are not independent. The following relations are established via integrations by part: $\gamma_{10} = -\gamma_{01}$, $\gamma_{21} = -\gamma_{12}$, $\gamma_{20} = -\gamma\gamma_{11}$.

The $\gamma_{ij}(r)$ are known function of r given by the moments of the two fields and their derivatives:

$$\gamma_{ij}(r) = \frac{1}{\sigma_i\sigma_j}\int {\rm d}k\,\exp(\imath kr)(\imath k)^i(-\imath k)^j P_s(k) \quad i\leqslant j\,, \tag{14}$$

with $P_s(k)$ the power spectrum of δ smoothed with a filter function (assumed Gaussian throughout this work). On expanding $\gamma_{ij}(r)$ at small separations $r \ll 1$ and substituting the spectral moments

$$\sigma_l^2 = \frac{1}{\pi}\int_0^\infty {\rm d}k\,k^{2l}P_s(k)\,, \tag{15}$$

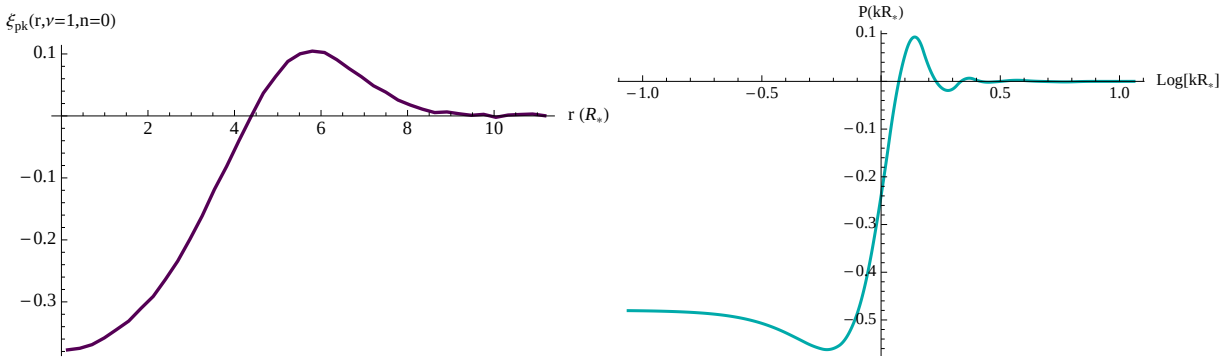


Figure 2. *Left panel*: Monte-Carlo estimation of the two-point correlation function of 1D peaks $\xi_{\rm pk}(r,\nu)$. The results is shown as a function of $r/R_\star$ assuming $n=0$ and a threshold $\nu=1$. *Right panel*: Corresponding power spectrum $P_{\rm pk}(k,\nu)$. The exclusion zone seen at short distance in the left panel makes the 1D peak power spectrum sub-Poissonian at small wavenumber (large separation).

it follows that

$$\gamma_{ij}(r)=\frac{(-1)^{l-i}}{\sigma_i\sigma_j}\sum_{k=0}^{\infty}(-1)^k\frac{\tilde{r}^{2k}}{(2k)!}\sigma^2_{l+k} \qquad (i+j=2l)$$

$$\gamma_{ij}(r)=\frac{(-1)^{1+l-i}}{\sigma_i\sigma_j}\sum_{k=0}^{\infty}(-1)^k\frac{\tilde{r}^{2k+1}}{(2k+1)!}\sigma^2_{l+k+1} \ (i+j=2l+1)$$

where $\tilde{r}=r/R$ is the separation in units of the smoothing length R. The determinant of the covariance matrix $\mathbf{C}$ is given at first order in the separation r by $\tilde{r}^{18}\times g(\{\sigma_i\}_{0\geqslant i\geqslant 5})$ where $g=\left(\sigma_2^6-\left(2\sigma_1^2\sigma_3^2+\sigma_0^2\sigma_4^2\right)\sigma_2^2+\sigma_0^2\sigma_3^4+\sigma_1^4\sigma_4^2\right)\times$ $\left(\sigma_3^6-\left(2\sigma_2^2\sigma_4^2+\sigma_1^2\sigma_5^2\right)\sigma_3^2+\sigma_1^2\sigma_4^4+\sigma_2^4\sigma_5^2\right)/74649600\sigma_0^4\sigma_1^4\sigma_2^4$ does not depend on the separation. Indeed, three eigenvalues of $\mathbf{C}$ are singular, respectively scaling like $\tilde{r}^{10}$, $\tilde{r}^6$ and $\tilde{r}^2$ and corresponding to the eigen-directions given by $(x-y)$, (x_1-y_1) and $(x_{11}-y_{11})$). This singularity proportional to $\tilde{r}^{-18}$ is the reason why the limit $\tilde{r}\to 0$ is difficult to handle numerically. Analytically it means that a series expansion to eighteenth order is needed for all terms.

3.1 Correlation of 1D peaks of same height

We evaluate the two-point correlation function of $D=1$ peaks upon applying the Monte-Carlo method described above to Eq.(10). Results are shown in the left panel of Fig.1 as a function of the spectral index n for a fixed peak height $\nu=1$. The exclusion zone shrinks to smaller separations and becomes more pronounced as n is increased because the addition of small scale power tends to sharpen the profile around local density maxima. In the right panel of Fig.1, we display $\xi_{\rm pk}(r,\nu)$ as a function of peak height for a white noise power spectrum $n=0$. For comparison, the dashed curves represent the two-point correlation of 1D signed critical points, which is obtained upon integrating equation (6) over the six field variables. Unsurprisingly, $\xi_{\rm crit}(r,\nu)$ matches $\xi_{\rm pk}(r,\nu)$ almost perfectly for prominent peaks ($\nu\gtrsim 2$) since, in this regime, a critical point is nearly always a local maximum.

Interestingly, the two-point correlation $\xi_{\rm crit}(r,\nu)$ is fully analytical regardless of the underlying density power spectrum. For the sake of readability however, we will not display its full expression here. However, we can take advantage of this analytic result, together with the fact that $\xi_{\rm pk}(r,\nu)$ agrees very well with its genus-like counterpart $\xi_{\rm crit}(r,\nu)$ at high threshold, to get insights into the short distance behaviour of the peak correlation function. The low-r limit of the two-point correlation function of signed critical points separated by r is given by

$$1+\xi_{\rm crit}(r,\nu)=\frac{e^{\frac{\left(2\gamma^2-1\right)\nu^2}{2\left(\gamma^2-1\right)}}\left(\gamma_{\#}^2\left(\gamma^2\gamma_*^4\left(\gamma^2+\nu^2-1\right)-2\gamma^2\gamma_*^2\left(\gamma^2+\nu^2-1\right)+\gamma^2\left(\nu^2+1\right)-1\right)+\left(\gamma^2-1\right)^2\right)}{12\gamma^3\left(1-\gamma^2\right)^{5/2}\nu^2\gamma_{\#}^2\sqrt{1-\gamma_*^2}\gamma_*^3}+\mathcal{O}(r) \qquad (16)$$

where we define the following shape parameters $\gamma=\sigma_1^2/\sigma_0/\sigma_2$, $\gamma_*=\sigma_2^2/\sigma_1/\sigma_3$ and $\gamma_{\#}=\sigma_3^2/\sigma_2/\sigma_4$.

For a power-law power spectrum $P_k(k)=Ak^n$, with spectral index $n>-1$, and a density field filtered with a Gaussian

kernel of radius R, the $\gamma_{ij}(r)$ are given by

$$\begin{aligned}\gamma_{00}(r) &= {}_1F_1\left(\frac{n+1}{2};\frac{1}{2};-\frac{\tilde{r}^2}{4}\right)\\ \gamma_{11}(r) &= {}_1F_1\left(\frac{n+3}{2};\frac{1}{2};-\frac{\tilde{r}^2}{4}\right)\\ \gamma_{22}(r) &= {}_1F_1\left(\frac{n+5}{2};\frac{1}{2};-\frac{\tilde{r}^2}{4}\right),\end{aligned} \tag{17}$$

and $\gamma_{02}(r) = -\gamma\gamma_{11}(r)$. Here, $\tilde{r}$ is the separation in unit of the smoothing length, ${}_1F_1$ is the Kummer confluent hypergeometric function and

$$\sigma_l^2 \propto R^{-1-n-2l}\,\Gamma\left(\frac{1}{2}+\frac{n}{2}+l\right)\ . \tag{18}$$

Therefore, the shape parameters are $\gamma = (1+n)/(3+n)$, $\gamma_* = (3+n)/(5+n)$ and $\gamma_\# = (5+n)/(7+n)$, and the determinant of C thus scales like $\gamma^2\left(2-\gamma^2\right)\left(1-\gamma^2\right)^{-3}\tilde{r}^{18}/18662400$, as advertised at the beginning of this Section. In the case where $n = 0$, the low-r behaviour of this function can be written as follows

$$1+\xi_{\rm crit}(r,\nu,n=0) = \frac{e^{\frac{\nu^2}{4}}\left(3\nu^2+8\right)}{8\sqrt{3}\nu^2} + \frac{e^{\frac{\nu^2}{4}}\left(128-15\nu^4\right)\tilde{r}^2}{1280\sqrt{3}\nu^2} + \frac{e^{\frac{\nu^2}{4}}\left(15\nu^4-64\right)\tilde{r}^4}{81920\sqrt{3}} + \mathcal{O}\left(\tilde{r}^5\right)\ , \tag{19}$$

which makes clear that $\xi_{\rm crit}$ diverges for $\nu = 0$. The dependence of the low-r expansion with the spectral index n for peaks of height $\nu = 1$ reads

$$1+\xi_{\rm crit}(r,\nu=1,n) = \frac{e^{\frac{1}{4}-\frac{n}{4}}(n+3)\left(n^2+4n+11\right)}{24(n+1)\sqrt{n^2+4n+3}} + \frac{e^{\frac{1}{4}-\frac{n}{4}}(n+3)\left(-25n^4+40n^3+410n^2+1464n+1695\right)\tilde{r}^2}{57600(n+1)\sqrt{n^2+4n+3}} + \mathcal{O}\left(\tilde{r}^3\right)\ , \tag{20}$$

which shows that the exclusion zone is more pronounced for high values of the spectral index. The same trend was also seen for peaks in the left panel of Fig.1. In general, the short distance behaviour of the correlation of critical points is consistent with a power-law in r, rather than the exponential suppression $\exp(-R_*^2/r^2)$ advocated in Lumsden et al. (1989).

As shown in Baldauf et al. (2013), the small-scale peak repulsion in the configuration space correlation function can have a significant impact on the power spectrum at small wavenumbers. To emphasize this point, Fig. 2 displays, in the left panel, the 1D peak correlation function as a function of $r/R_\star$ assuming $n = 0$ and a threshold $\nu = 1$ and, in the right panel, the corresponding power spectrum obtained by a simple Fourier transform of the real space Monte-Carlo result. In this particular case, the power spectrum is approximately white for all wavenumbers $k \lesssim R_*$, with $P_{\rm pk}(k \lesssim R_*, \nu = 1) \approx -0.5$.

3.2 How strong is the small-scale exclusion ?

So far, we have assumed that the peaks under consideration have exactly the same height. This is clearly a very special case, since a realistic sample of haloes is likely to be made up by a range of masses or smoothing scales and thus of different peak heights ν. As we have seen in Fig. 1, the exclusion region is often reduced in the sense that $1+\xi_{\rm pk}(r,\nu)$ does not reach zero at the origin. In order to ascertain whether this is a robust feature, we have also computed the correlation function of 1D peaks and critical points with different heights $\nu_1 = 1-\Delta\nu/2$ and $\nu_2 = 1+\Delta\nu/2$, where $\Delta\nu > 0$ is the height difference. For the critical points we have at small separations

$$\begin{aligned}1+\xi_{\rm crit}(r) = &\left[\frac{23328\sqrt{3}\Delta\nu^2R_\star^8}{r^8\left(\Delta\nu^2-4\nu^2\right)} + \frac{1296\sqrt{3}\Delta\nu^2R_\star^6}{r^6\left(\Delta\nu^2-4\nu^2\right)} + \frac{72\sqrt{3}\Delta\nu^2R_\star^4}{r^4\left(\Delta\nu^2-4\nu^2\right)} + \frac{24\sqrt{3}\Delta\nu^2R_\star^2}{5r^2\left(\Delta\nu^2-4\nu^2\right)}\right.\\ &\left.-\frac{16\Delta\nu^2+15\nu^2+40}{10\sqrt{3}\left(\Delta\nu^2-4\nu^2\right)}\right]\exp\left[\frac{7\Delta\nu^2}{80}+\frac{\nu^2}{4}-\frac{324\Delta\nu^2R_\star^6}{r^6}+\frac{27\Delta\nu^2R_\star^4}{r^4}-\frac{9\Delta\nu^2R_\star^2}{5r^2}\right]\end{aligned} \tag{21}$$

Comparing this expression to Eq. (19) above, we see that for unequal heights, inverse powers of separation arise. The r^{-6} term in the exponential drives the correlation function to -1 on very large scales. We can estimate where the leading inverse power in the exponent causes a one percent correction to $1+\xi$ from $r_{1\%} \approx (180\Delta\nu)^{1/3}\,R_\star$. Numerical results for peaks are shown in Fig. 3 for different choices of $\Delta\nu$. While the correlation $1+\xi_{\rm pk}(r,\nu_1,\nu_2)$ of peaks of different heights tends towards zero at small separation, it converges towards a finite non-zero value when the peaks have exactly the same height. In Fig. 4, we compare the exclusion of peaks and critical points to the approximation of Eq. (21). The scales where the finite separation results deviate from the equal height case are approximately the same for peaks and critical points and $r_{1\%}$ is a good indicator of this scale. Therefore, the behaviour of $\xi_{\rm pk}$ in the limit $r \to 0$ strongly depends on the peak height difference. This can be easily understood as follows: consider two peaks infinitesimally close to each other. Very stringent constraints on the first and second derivatives of the density field are thus required to bridge them. Clearly, the constraints will be more draconian the larger the height difference. Therefore, this configuration becomes increasingly unlikely with increasing $\Delta\nu > 0$, and thus

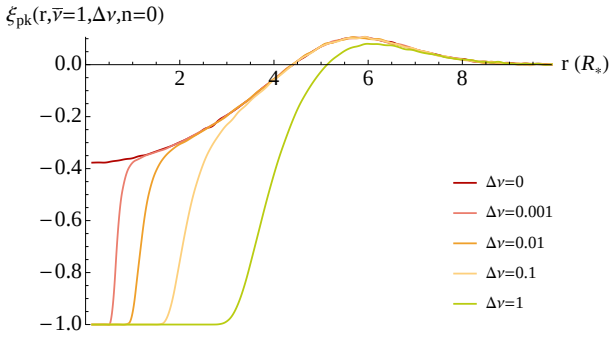


Figure 3. The two-point correlation function of 1D peaks with height $\nu_1 = 1 - \Delta\nu/2$ and $\nu_2 = 1 + \Delta\nu/2$ is shown as a function of the height difference $\Delta\nu$ as labeled in the Figure. A fixed value of $\bar{\nu} \equiv (\nu_1 + \nu_2)/2 = 1$ was assumed. All the correlation functions were estimated using the Monte-Carlo method described in Sec. 2.

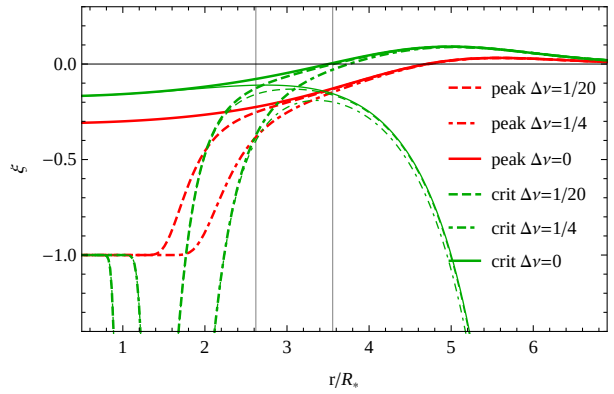


Figure 4. Correlation function of peaks (red) and critical points (green) for $\bar{\nu} = 3/2$ and for $\Delta\nu = 0$ (solid), $\Delta\nu = 1/10$ (dashed) and $\Delta\nu = 1/4$ (dot-dashed). The scales, where the $\Delta\nu = 0$ and $\Delta\nu \neq 0$ cases deviate are similar for peaks and critical points. For the critical points we overplot the low-r expansion Eq. (21). This expansion provides a good description of the low-r behaviour up to 2-3 $R_\star$. The vertical lines indicate the respective $r_{1\%}$ scales, and provide a useful estimate of the scale where the $\Delta\nu = 0$ and $\Delta\nu \neq 0$ cases start to deviate in the full calculation.

$1 + \xi_{\rm pk}(r, \nu_1, \nu_2)$ rapidly drops to zero as the separation decreases. The same behaviour is, of course, expected to hold for peaks with different smoothing scales and for integrals over bins in peak height.

4 LARGE SCALES: PERTURBATIVE BIAS EXPANSION

Perturbative bias expansions have been widely used to predict clustering statistics of dark matter halos and galaxies. However, no study so far has explored the convergence properties of these series because of the highly non-linear (non-perturbative) effects induced by small-scale exclusion. In this Section, we wish to address this issue using the clustering of 1D peaks as a proxy for the two-halo term.

4.1 Methodology

As was shown in Desjacques (2013), the two-point correlation function of 3D peaks can be computed from an perturbative, local bias expansion in which the coefficients (bias parameters) are computed from a generalized peak-background split Ansatz. This procedure is fairly general and it applies to any 'point' process of a Gaussian (and possibly non-Gaussian) random field. Therefore, it should certainly describe the two-point function of our 1D peak. In 1D, the perturbative bias expansion is constructed from 3 rotationally invariant quantities, i.e. $x = \rho/\sigma_0$, $\eta^2 = x_1^2$ and $u = -\nabla_i^2\rho/\sigma_2$. The "localized" number density of peaks of height ν then is

$$n_{\rm pk}(\mathbf{X}) = \frac{\sigma_2}{\sigma_1}|u|\delta_D(\eta)\delta_D(x-\nu) \tag{22}$$

The knowledge of $n_{\rm pk}(\nu)$ suffices to derive the bias parameters associated with this point process at all orders. Namely, the probability density for the variables $\mathbf{X} = (x, u, \eta^2)$ is the product of a bivariate normal $\mathcal{N}(x, u)$ with a chi-square distribution $\chi_1^2(\eta^2)$ with one degree of freedom. To construct the perturbative bias expansion, we proceed as in Desjacques (2013) and

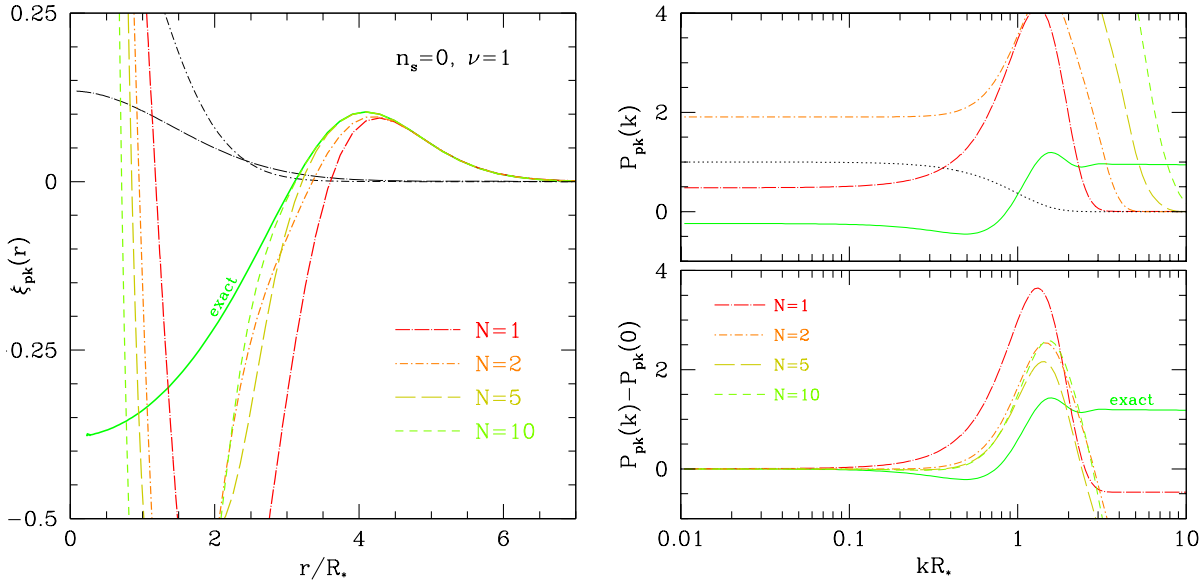


Figure 5. *Left panel*: Comparison between the exact correlation function of 1D peaks and perturbative series obtained upon truncating the peak bias expansion Eq.(23) at order $N = 1$, 2, 5 and 10. For illustration, the black curves show linear and quadratic local bias approximations and perform significantly worse than the peak bias expansion. A white noise power spectrum with Gaussian filtering and a peak height $\nu = 1$ were assumed. *Top right*: Exact peak power spectrum (solid) compared to various approximations obtained upon truncating the bias expansion at order $N = 1$, 2, 5 and 10. The solid curve is the density power spectrum. *Bottom right*: same as top panel but the zero-lag value $P_{\rm pk}(0)$ has been subtracted off from the power spectra. In all cases, a white noise power spectrum with Gaussian smoothing and a peak height of unity were assumed.

perturb the central values of the distribution (the peak-background split). This perturbation can be represented as a series in the appropriate orthogonal polynomials, i.e. Hermite polynomials for $\mathcal{N}(x,u)$ and generalized Laguerre polynomials for $\chi^2_1(\eta^2)$. The perturbative bias expansion describing 1D peaks thus is

$$\begin{aligned}\delta_{\rm pk}(r) &= \sum_{i,j,k} \frac{\sigma_0^i \sigma_2^j \sigma_1^{2k}}{i!j!} \frac{\Gamma(1/2)}{\Gamma(k+1/2)} b_{ij}\chi_k x^i(r) u^j(r) \eta^{2k}(r) \\ &= \sigma_0 b_{10} x(r) + \sigma_2 b_{01} u(r) + \frac{1}{2}\sigma_0^2 b_{20} x^2(r) + \sigma_0\sigma_2 b_{11} x(r)u(r) + \frac{1}{2}\sigma_2^2 b_{02} u^2(r) + \sigma_1^2\chi_1\eta^2(r) + \dots \end{aligned} \tag{23}$$

where r is the 1D comoving coordinate. The bias parameters are the ensemble averages

$$\sigma_0^i\sigma_2^j b_{ij} = \frac{1}{\bar{n}_{\rm pk}(\nu)} \int {\rm d}\mathbf{X}\, n_{\rm pk}(\mathbf{X})\mathcal{N}(\mathbf{X})\, H_{ij}(x,u) \tag{24}$$

$$\sigma_1^{2k}\chi_k = \frac{(-1)^k}{\bar{n}_{\rm pk}(\nu)} \int {\rm d}\mathbf{X}\, n_{\rm pk}(\mathbf{X})\mathcal{N}(\mathbf{X})\, L_k^{(-1/2)}(\eta^2/2)\ . \tag{25}$$

The latter evaluates to $\chi_k = (-1/2)^k(2k-1)!!/(k!\sigma_1^{2k})$. The average peak number density is given by

$$\begin{aligned}\bar{n}_{\rm pk}(\nu) &= \frac{\sigma_2}{\sigma_1}\int {\rm d}\mathbf{X}\mathcal{N}(\mathbf{X})\,|u|\delta_D(\eta)\delta_D(x-\nu) \\ &= \frac{\sigma_2}{\sigma_1}\frac{1}{\sqrt{2\pi}}\int {\rm d}u\,|u|\,\mathcal{N}(\nu,u) \\ &= \frac{\sigma_2}{\sigma_1}\left\{\frac{1}{2}\gamma\nu\left[1+{\rm Erf}\left(\frac{\gamma\nu}{\sqrt{2-2\gamma^2}}\right)\right] + \sqrt{\frac{1-\gamma^2}{2\pi}}e^{-\frac{\gamma^2\nu^2}{2(1-\gamma^2)}}\right\}\frac{e^{-\nu^2/2}}{2\pi} \\ &\equiv \frac{1}{2\pi R_\star}G(\gamma,\gamma\nu)e^{-\nu^2/2}\ , \end{aligned} \tag{26}$$

where $R_\star = \sigma_1/\sigma_2$. Note that, for these expressions to make sense, we must define $\delta_D(\eta^2) = \delta_D(x_1)/(2x_1)$.

In 1D, we could also use x_1 rather than $\eta^2 = x_1$ as independent variable. x_1 is normally distributed, so that the joint probability distribution reads $P_1(\mathbf{y}) = \mathcal{N}(x,u)\mathcal{N}(x_1)$, and the bias parameters associated with x_1 are

$$\begin{aligned}\sigma_1^k\chi_k &= \frac{1}{\bar{n}_{\rm pk}}\int {\rm d}\mathbf{X}\mathcal{N}(\mathbf{X})\, H_k(x_1) \\ &= (-1)^{k/2}\,(k-1)!! \qquad \text{if } k \text{ is even}\ . \end{aligned} \tag{27}$$

In both cases, we recover the same series expansion. Namely, on perturbing either $\mathcal{N}(x_1)$ or $\chi^2_1(\eta^2)$ with a long-wavelength perturbation such that $x_1 \to x_1 + \eta_l$ and $\eta^2 \to \eta^2 + \eta_l^2$, we find Desjacques (2013)

$$\text{i)} \qquad \sum_{n=0}^{\infty} \frac{1}{n!}(\sigma_1^n \chi_n)\eta_l^n = \sum_{k=0}^{\infty} \frac{1}{(2k)!}(-1)^k (2k-1)!!\, \eta_l^{2k} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!}\eta_l^{2k} \quad \text{(Hermite)}$$

$$\text{ii)} \qquad \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+k+1)}(\sigma_1^{2k}\chi_k)\left(\frac{\eta_l^2}{2}\right)^k = \sum_{k=0}^{\infty} \frac{2^k}{(2k-1)!!}(-1/2)^k \frac{(2k-1)!!}{k!}\left(\frac{\eta_l^2}{2}\right)^k = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!}\eta_l^{2k} \quad \text{(Laguerre)}\ .$$

In other words, we could equally work with the bias factors χ_k expressed either as Hermite or Laguerre. We could also have followed Gay et al. (2012) and works with the variables $(z = (x - \gamma u)/\sqrt{1-\gamma^2}, u)$ which have the advantage of being statistically independent. In what follows, we will stick to the variable η^2.

The 1D peak two-point correlation $\xi_{\rm pk}(r)$ can be perturbatively computed from Eq.(23). In practice, one evaluates the ensemble average $\langle \delta_{\rm pk}(r_1)\delta_{\rm pk}(r_2)\rangle$, ensuring to discard all the terms involving zero-lag moments. At first order, this is

$$\xi_{\rm pk}(r) = b_{10}^2 \xi_0(r) + 2 b_{10} b_{01} \xi_2(r) + b_{01}^2 \xi_4(r)\ , \tag{28}$$

where

$$\xi_l(r) = \frac{1}{\pi}\int_0^\infty dk\, k^l P(k) \times \begin{cases} \cos(kr) & l \text{ even} \\ \sin(kr) & l \text{ odd} \end{cases}\ . \tag{29}$$

Note that $\xi_{2i}(r) = \gamma_{ii}(r)$, where $\gamma_{ij}(r)$ is defined in Eq.(14). For Gaussian initial conditions (which we assume throughout this paper), the Nth order contribution to $\xi_{\rm pk}(r)$ involves $N(N+1)/2$ distinct combinations of x and u correlators together with N^2 terms involving correlators of η^2. Therefore, the number of terms scales like N^2. Consequently, the 1D peak correlation up to nth order involves $\mathcal{O}(N^3)$. All these contributions can be expressed as a product of the 6 possible connected two-point correlators $\langle\nu_1\nu_2\rangle$, $\langle\nu_1\eta_2\rangle$ etc. (where the subscripts denote the positions r_1 and r_2). However, the coefficients are product of Nth order bias parameters and change from term to term. Therefore, the bias perturbative expansion will be of limited use unless the bias coefficients can be computed quickly. In practice, exploiting the recurrence among the orthogonal polynomials can help reducing the computational cost.

4.2 Convergence in real and Fourier space

Armed with these results, we can assess the convergence properties of the 1D peak perturbative bias expansion. For illustrative purposes, we will consider a power-law power spectrum $P(k) \propto k^n$ with $n = 0$ and a peak height $\nu = 1$. In the left panel of Fig. 5, we compare the exact result)solid green curve) with the $N = 1$, 2, 5 and 10th order perturbative approximations (in color). While the latter capture the excess correlation at a few $R_\star$ relatively well, the convergence to the exact result is fairly slow at shorter separations where exclusion effects become important. In any case, the peak bias expansion Eq.(23) perform significantly better than a "local bias" approximation, in which only the dependence on $x(r)$ is retained in the perturbative series Eq.(23). The first and second-order "local bias" approximations are shown as the black curves, and clearly furnish a poor fit to the exact result for $r/R_\star \lesssim$ a few.

The upper right panel of Fig.5 displays the resulting power spectra obtained by taking the Fourier transform of $\xi_{\rm pk}(r)$ and its various perturbative approximations. For sake of comparison, the dotted black curve represents the first order approximation. Small-scale exclusion translates into a white-noise contribution in the limit $k \to 0$ which makes the shot-noise non-Poissonian, in agreement with the findings of Smith et al. (2007); Baldauf et al. (2013). Note that such a k^0 tail also arises in the clustering of thresholded regions Beltrán Jiménez & Durrer (2011). The magnitude of the white-noise correction changes with the order N of the approximation because it receives contributions at all orders. As shown in the previous Section, deriving an analytic expression valid throughout the exclusion zone is practically impossible and, therefore, there is no hope to obtain exact expression for these deviations from Poisson noise. The bottom right panel of Fig.5 displays the power spectra once the white noise correction $P(k = 0)$ (which generally depends on the order N of the approximation) has been subtracted. The relatively slow convergence of the perturbative approximations to the 1D peak power spectrum towards the exact result reflects the behaviour seen in configuration space.

This has implications for the convergence of perturbative bias expansions of actual dark matter halos.

5 ZEL'DOVICH DISPLACEMENT OF PEAKS

So far we have been concerned with peaks in Lagrangian space, but haloes are observed in Eulerian space, i.e. after gravitational evolution. Let us use the Lagrangian displacement Ψ to relate the Lagrangian peak position q to its final Eulerian position x

$$x_{\rm pk} = q_{\rm pk} + \zeta \Psi_{\rm pk}(q_{\rm pk})\ , \tag{30}$$

where ζ is the linear growth factor and the displacement satisfies the equation of motion

$$\ddot{x} + H\dot{x} = -\nabla_x \phi \,. \tag{31}$$

The linear solution to this equation in three dimensions is known as the Zel'dovich approximation and describes ballistic evolution following according to the initial velocity. We will assume that the peak displacement is related to the mean displacement of the peak patch, i.e., by $\Psi(k) = -\imath k/k^2 \delta(k) W_R(k)$, where $W_R(k)$ is the smoothing window. Conveniently, in one dimension, the Zel'dovich solution is the exact solution McQuinn & White (2015).[1]

We can now write the Eulerian halo/peak overdensity as a sum over Eulerian peak positions, which are in turn related to their respective Lagrangian proto-halo positions

$$1 + \delta_{\rm h}(x) = \frac{1}{\bar{n}_{\rm pk}} \sum_{\rm h} \delta_{\rm D}\left(x - x_{\rm h}\right) = \int {\rm d}q' \delta_{\rm D}\left[x - q' - \Psi(q')\right] \sum_{\rm pk} \delta_{\rm D}(q' - q_{\rm pk}) \tag{34}$$

$$= \int {\rm d}q' \int \frac{{\rm d}k}{2\pi} \exp\left[\imath k (x - q' - \Psi(q'))\right] \delta_{\rm D}(q' - q_{\rm pk}) \tag{35}$$

$$= \int {\rm d}q' \int \frac{{\rm d}k}{2\pi} \exp\left[\imath k (x - q')\right] \delta_{\rm D}(q' - q_{\rm pk}) \exp\left[-\imath k \Psi(q')\right] \tag{36}$$

$$= \int {\rm d}q' \int \frac{{\rm d}k}{2\pi} \exp\left[\imath k (x - q')\right] |\delta''(q)| \delta_{\rm D}\left[\delta'(q')\right] \exp\left[-\imath k \Psi(q')\right] \tag{37}$$

For the correlation of evolved peaks in one dimension we obtain

$$\xi(r) = \langle \delta_{\rm h}(x) \delta_{\rm h}(x+r) \rangle = \frac{1}{\bar{n}_{\rm pk}^2} \int {\rm d}Q \int \frac{{\rm d}k}{2\pi} \int \exp\left[\imath k (Q - r)\right]$$
$$\times \ \langle \exp\left[-\imath k \zeta (\Psi_1 - \Psi_2)\right] |\delta''(q_1)||\delta''(q_2)| \delta_{\rm D}\left[\delta'(q_1)\right] \delta_{\rm D}\left[\delta'(q_2)\right] \rangle - 1 \tag{38}$$

where $Q = q_2 - q_1$ is the Lagrangian separation of the peaks and Ψ_1 and Ψ_2 are the displacements at the respective positions. The displacement field is an additional stochastic variable that needs to be averaged over under the peak constraint. For this purpose we need to append the covariance matrix by components that describe the auto-covariance of the displacement and the cross-covariance between displacement and the density field and its derivatives.
Splitting the state vector into $\mathbf{X} = (\mathbf{m}, \mathbf{n})$ where $\mathbf{m} = (\sigma_0 \nu_1, 0, -\sigma_2 u_1, \sigma_0 \nu_2, 0, -\sigma_2 u_2)$ and $\mathbf{n} = (\Psi_1, \Psi_2)$, the covariance matrix takes the following schematic form after shifting the displacement variables to the lower right corner

$$M = \begin{pmatrix} \langle \mathbf{m}, \mathbf{m} \rangle & \langle \mathbf{m}, \mathbf{n} \rangle \\ \langle \mathbf{n}, \mathbf{m} \rangle & \langle \mathbf{n}, \mathbf{n} \rangle \end{pmatrix} = \begin{pmatrix} A & B \\ B^{\rm T} & C \end{pmatrix} = \begin{pmatrix} S & T \\ T^{\rm T} & V \end{pmatrix}^{-1} \tag{39}$$

and for the determinant we have $\det M = \det A \, \det V$ with $V = (C - B^{\rm T} A^{-1} B)^{-1}$.
We can now write $\imath k (\Psi_1 - \Psi_2)$ as $\imath k p^T n$ with $p^T = (1, -1)$. The exponential arising from the PDF and the shift can be written as

$$\mathbf{Y}^T M^{-1} \mathbf{Y} + 2\imath k \mathbf{p}^T \mathbf{n} = \mathbf{m}^T S \mathbf{m} + 2\mathbf{m}^T T \mathbf{n} + \mathbf{n}^T V \mathbf{n} + 2\imath\zeta k \mathbf{p}^T \mathbf{n}$$
$$= \mathbf{m}^T S \mathbf{m} + 2\mathbf{m}^T T \mathbf{n} + 2\mu^T V \mathbf{n} + 2\imath\zeta k \mathbf{p}^T \mathbf{n} - \mu^T V \mu + (\mathbf{n} - \mu)^T V (\mathbf{n} - \mu) \tag{40}$$

where we completed the square to isolate the displacement part of the PDF. Thus we have with $\mu^T = -\mathbf{m}^T T - \imath k \mathbf{p}^T$

$$\mathbf{Y}^T M^{-1} \mathbf{Y} + 2\imath k \mathbf{p}^T \mathbf{n} = \mathbf{m}^T A^{-1} \mathbf{m} + \zeta^2 k^2 \mathbf{p}^T V^{-1} \mathbf{p} - 2\imath\zeta k \mathbf{m}^T A^{-1} B \mathbf{p} + (\mathbf{n} - \mu)^T V (\mathbf{n} - \mu) \tag{41}$$

The vector of displacements $\mathbf{n}$ can be integrated out, yielding unity and leaves us with

$$\langle \delta_{\rm h}(x) \delta_{\rm h}(x+r) \rangle = \frac{1}{\bar{n}_{\rm pk}^2} \int {\rm d}Q \int \frac{{\rm d}k}{2\pi} \int {\rm d}\mathbf{m} \exp\left[\imath k (Q - r)\right]$$
$$\frac{1}{\sqrt{(2\pi)^6 \det A}} \exp\left[-\frac{1}{2}\mathbf{m}^T A^{-1} \mathbf{m} - \frac{1}{2}\zeta^2 k^2 \mathbf{p}^T V^{-1} \mathbf{p} - \imath\zeta k \mathbf{m}^T A^{-1} B \mathbf{p}\right] - 1 \tag{42}$$

[1] From $[1 + \delta(x)] {\rm d}x = {\rm d}q$ we have $\delta(x) = 1/J - 1$ with $J = |1 + \nabla_q \Psi|$. Then we can take the Eulerian divergence of the equation of motion

$$\nabla_x \left[\ddot{\Psi} + H\dot{\Psi}\right] = \frac{3}{2}\mathcal{H}^2 \delta \tag{32}$$

Rewriting the Eulerian as a Lagrangian derivative $\nabla_x = (1 + \nabla_q \Psi)^{-1} \nabla_q$ we finally have

$$\nabla_q \left[\ddot{\Psi} + H\dot{\Psi}\right] = \frac{3}{2}\mathcal{H}^2 \nabla_q \Psi \tag{33}$$

which is a linear equation for Ψ that can be solved exactly by $\Psi = \zeta \Psi_0$ before shell crossing, where ζ is the linear growth.

Fourier transforming and expanding this expression for small correlations or (i.e. $Q \to \infty$) we have

$$P(k) = \frac{1}{\bar{n}_{\rm pk}^2}\int \mathrm{d}Q \int \mathrm{d}\mathbf{m} \exp\left[\imath k Q\right] \Big\{ b_{10}(\nu_1,q_1)b_{10}(\nu_2,q_2)\xi_0(Q) + \zeta^2 k^2 \xi_{vv}(Q) - \imath\zeta k \big[b_{10}(\nu_1,q_1)+b_{10}(\nu_2,q_2)\big]\xi_{v0}(Q)\Big\}$$
$$\times \exp\left[-k^2\sigma^2_{v,\rm pk}\right]\prod_i \exp\left[-\frac{1}{2}\frac{(q_i-\gamma\nu_i)^2}{1-\gamma^2} - \frac{1}{2}\nu_i^2\right] \tag{43}$$

where we used

$$\mathbf{p}^T V \mathbf{p} \approx 2\sigma^2_{v,\rm pk} - 2\left(\xi_{vv} - 2\xi_{v1}\frac{\sigma^2_{v1}}{\sigma^2_1} + \xi_1 \frac{\sigma^4_{v1}}{\sigma^4_1}\right) \tag{44}$$

$$\mathbf{m}^T A^{-1} B \mathbf{p} \approx \frac{1}{\sigma_0}\left(\frac{\nu_1-\gamma q_1}{1-\gamma^2} + \frac{\nu_2-\gamma q_2}{1-\gamma^2}\right)\xi_{v0} + \dots \tag{45}$$

and defined the peak velocity/displacement dispersion

$$\sigma^2_{v,\rm pk} = \sigma^2_{-1} - \frac{\sigma^4_{v1}}{\sigma^2_1} \tag{46}$$

Performing the Q integral and the average over peak curvature we have at low wavenumbers and ignoring shotnoise contributions

$$P(k) \approx \big[b_{10}(\nu_1)+\zeta\big]\big[b_{10}(\nu_2)+\zeta\big]P(k), \tag{47}$$

i.e., the large scale bias is given by $b_{10}+\zeta$ as one would have expected from Mo & White (1996).
Let us come back to the full expression Eq. (42). Together with the pre factor $\exp\left[\imath k(r-Q)\right]$, this is a Gaussian integral in k, which is readily performed to yield

$$\langle \delta_{\rm h}(x)\delta_{\rm h}(x+r)\rangle = \frac{1}{\bar{n}_{\rm pk}^2}\int \mathrm{d}Q \int \mathrm{d}\mathbf{m}\ \frac{1}{\sqrt{(2\pi)^6 \det A}}\exp\left[-\frac{1}{2}\mathbf{m}^T A^{-1}\mathbf{m}\right]$$
$$\frac{1}{\sqrt{2\pi\zeta^2 \mathbf{p}^T V^{-1}\mathbf{p}}}\exp\left[-\frac{1}{2}\frac{(r-Q-\zeta\mathbf{m}^T A^{-1}B\mathbf{p})^2}{\zeta^2\mathbf{p}^T V^{-1}\mathbf{p}}\right] - 1 \tag{48}$$
$$= \int \mathrm{d}Q\ F(Q|r) - 1 \tag{49}$$

This is a convolution between the Eulerian and Lagrangian distances r and Q. For fixed Eulerian distance the integral is approximately Gaussian peaked at $Q = r - \zeta\mathbf{m}^T A^{-1}B\mathbf{p}$ (see Fig. 6). As the growth factor ζ goes to zero the last part of the above integral becomes a Dirac delta function for $Q-r$ and we recover the Lagrangian expression. The shift term $\zeta\mathbf{m}^T A^{-1}B\mathbf{p}$ is negligible on large scales but can lead to significant corrections on small scales and actually fills the $\xi = -1$ region on small scales.
For the calculation of the effect of displacements we have to resort to one dimensional power spectra that grow steeper than k^0 for low wavenumbers since otherwise the velocity correlators diverge. In particular, we need a power spectrum that has a slope $n > 2$ and we will choose $n = 2$ for definiteness. We show the evolved correlation function and bias in Fig. 7.

In Baldauf et al. (2013) it has been shown in simulations that the stochasticity amplitude on large scales is the same in Lagrangian and Eulerian space. On small scales the corrections to the fiducial $1/\bar{n}$ stochasticity have to vanish, and they do so for haloes in Eulerian and proto-haloes in Lagrangian space. The only change in the behaviour is that the transition happens at higher wavenumbers in Eulerian space. We would like to explore to what extend our 1D peak model can reproduce this behaviour. In the left panel of Fig. 8 we show the power spectrum of peaks in Lagrangian space and their stochasticity estimated as $P_{\rm pk,pk} - b_{10}^2 P$. We see that the large scale amplitude is the same for initial and evolved haloes. After correction for the large scale bias in Eulerian space, we see that the transition between the non-zero and zero stochasticity correction regimes is pushed to higher wavenumbers.

The mean motion of peaks can be understood based on their initial velocity statistics. Let us for this purpose consider the mean infall $\langle (v_1-v_2)\delta_1\delta_2\rangle$. We see that the mean infall of haloes deviates from the underlying dark matter motion quite significantly. In a fashion similar to what was presented above for displaced peaks and with $\mathbf{p} = (1,-1)$ we have

$$v_{12} := \langle (v_2-v_1)(1+\delta_1)(1+\delta_2)\rangle = \frac{1}{\bar{n}_{\rm pk}^2}\int \mathrm{d}\mathbf{m}\ \mathbf{m}^T T\mathbf{p}\ \frac{1}{\sqrt{(2\pi)^6\det A}}\exp\left[-\frac{1}{2}\mathbf{m}^T A^{-1}\mathbf{m}\right]\ . \tag{50}$$

On large scales this quantity can be approximated by its linear (scale dependent) bias expansion

$$v_{12} \approx 2b_{10}\left(\xi_{v0} + \frac{\sigma^2_{v1}}{\sigma^2_1}\xi_{12}\right) + b_{01}\left(-\xi_{v2} + \frac{\sigma^2_{v1}}{\sigma^2_1}\xi_{32}\right)\ , \tag{51}$$

whereas the local bias model yields $v_{12} \approx 2b_{10}\xi_{v0}$. In Fig. 8 we show the mean relative velocity of the same sample of peaks considered for the evolution above, i.e., peaks of height $\bar{\nu} = 3/2$ and $\Delta\nu = 3/5$ in a $n = 3$ power law density field. For the underlying matter ditribution we have $v_{12} \approx -2\langle v\delta\rangle$. We clearly see that matter and peak mean streaming differ for scales

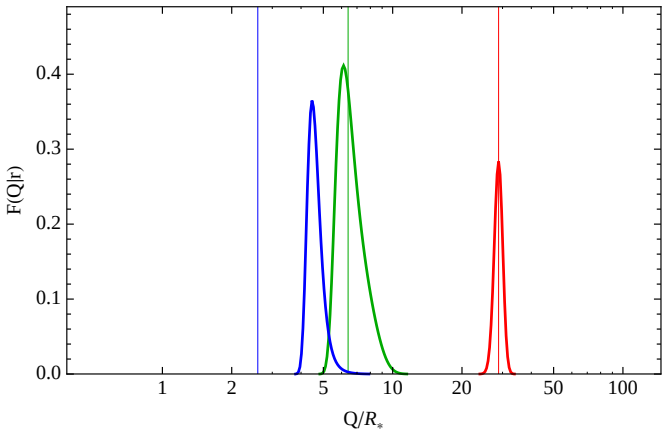


Figure 6. Contribution to the integral over Lagrangian separations in Eq. (49). The vertical lines indicate the Eulerian separation r and the corresponding curves indicate the support of the integral. We see that for large scales, the evolution merely corresponds to a convolution with a Gaussian, whereas for small scales there is a an offset in the support to larger scales and the convolution kernel is skewed.

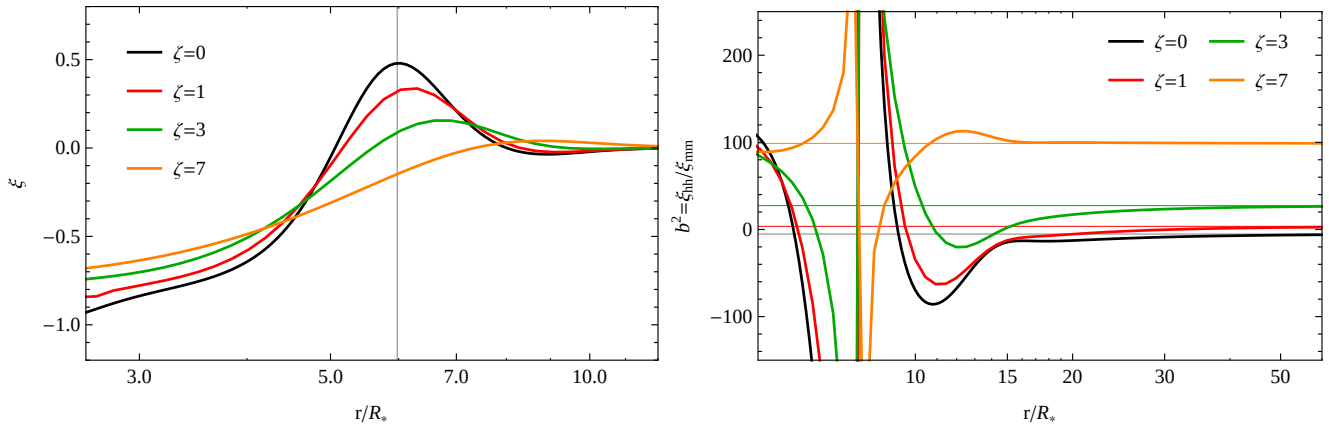


Figure 7. *Left panel:* Small scale correlation function of inital and evolved peaks in a $n = 3$ power law density field. We have chosen peaks of height $\bar{\nu} = 6/5$ and $\Delta\nu = 3/5$ to highlight the behaviour in presence of exclusion. The linear growth factor increases from the initial conditions $\zeta = 0$ to some arbitrary final time $\zeta = 7$. *Right panel:* Linear bias of the peak correlation functions with respect to the linear matter correlation function. We clearly see that the bias asymptotes to the expected $b_{10} + \zeta$ behaviour for large separations indicated by the horizontal lines.

$r < 10R_\star$. This deviation is captured by the scale dependent peak velocity bias Eq (51) down to $r \approx 7R_\star$. Thus, the linear peak velocity bias has starts to deviate at larger separations than the linear peak density bias.

6 CONCLUSIONS

Halo exclusion is an essential ingredient towards a realistic description of the two-halo term, which encodes most of the cosmological information that can be extracted from the two-point correlation of biased tracers. Halo exclusion stems from the fact that one cannot find two peaks of the mass density field (halos) arbitrary close to each other or even overlap. Until very recently however, this effect had been either completely ignored or crudely modeled by setting $\xi_{\rm h}(r) = -1$ at separations r less than the sum of the halo virial radii, while sticking to the linear bias approximation at larger distances.

In this paper, we have investigated this small-scale exclusion using a simple, well-motivated approximation: the clustering of Gaussian density peaks and, more general, critical points in one dimension. After studying the small-r behaviour for various power-law spectra and its sensitivity to the peak height, we have shown that the two-point correlation function of 1D density peaks differs significantly from the crude, aforementioned prescription. We have also explored how peak exclusion affects the convergence of the perturbative bias expansion in real and Fourier space. Finally, we have included the displacement

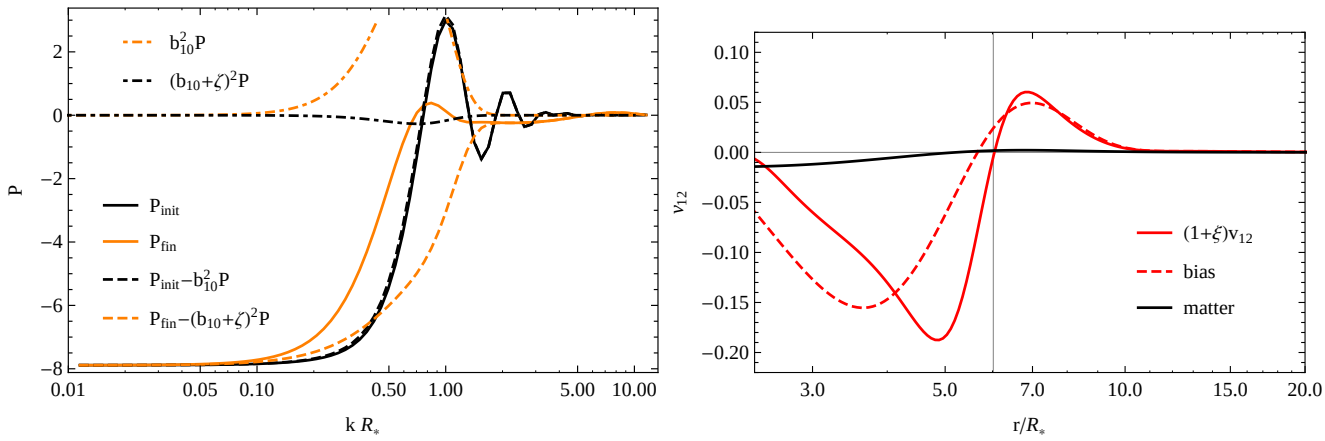


Figure 8. *Left panel:* Evolution of the peak-peak power spectrum and its noise component $P_{\rm pk,pk} - b^2 P$ from the initial conditions to the final configuration at $\zeta = 7$. While the amplitude remains constant on large scales, the infall causes *Right panel:* Mean relative velocity v_{12} of 1D peaks (red solid). The peak velocities clearly deviate from the velocities of random dark matter particles (orange solid) on small scales. On large scales this deviation can be described by a linear velocity bias (red dashed), which however fails to describe the non-perturbative behaviour in the exclusion regime. We see that below $r = 6R_\star$ peaks are moving towards each other and away from each other for larger separations. The vertical gray line indicates the peak of the correlation function in Fig. 7. Note that on top of this mean flow, there is also a velocity dispersion.

from the initial to final peak position according to the Zel'dovich approximation to clarify how exclusion effects mix up with scale-dependencies induced by the nonlinear gravitational evolution.

Our key findings can be summarized in the following points:

(i) the correlation function of equal height peaks or critical points ($\Delta\nu = 0$) asymptotes to a finite non-zero number at small scales

(ii) the correlation function of unequal height peaks or critical points ($\Delta\nu \neq 0$) deviates from the aforementioned case on small scales and asymptotes to exactly -1, where scale of deviation from (i) scales with the difference in height as $r \propto \Delta\nu^{1/3}$

(iii) the local bias expansion fails to describe the scale dependence starting at very large scales, the peak bias series including derivative operators fares better but its convergence in the $r \approx R_\star$-regime is very slow and it completely fails in the exclusion regime

(iv) time evolution enhances the large scale clustering according to the well known $b_{10}+\zeta$ behaviour and leads to non-trivial modifications in the exclusion regime

Even though our findings apply, strictly speaking, to the clustering of tracers in one-dimensional density fields, they provide useful insights into halo exclusion and its impact on the two-halo term in a realistic setting in three dimensions. It would certainly be interesting but computationally more challenging to study these effects for peaks in three dimensional density fields.

ACKNOWLEDGMENTS

We thank D. Pogosyan for useful comments during the course of this work. VD thanks the LABEX "Institut de Lagrange de Paris" for funding and acknowledges support from the Swiss National Science Foundation. TB gratefully acknowledges support by the Institute for Advanced Study through the Corning Glass works foundation fund. CP and SC thank the University of Geneva for funding and the community of `http://mathematica.stackexchange.com` for technical advice. This research is part of the Spin(e) (ANR-13-BS05-0005, `http://cosmicorigin.org`) and the Cosmo@NLO grants of the French Agence Nationale de la Recherche.

References

Baldauf T., Seljak U., Smith R. E., Hamaus N., Desjacques V., 2013, Phys. Rev. D, 88, 083507
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, Astrophys. J., 304, 15
Beltrán Jiménez J., Durrer R., 2011, Phys. Rev. D, 83, 103509
Casas-Miranda R., Mo H. J., Sheth R. K., Boerner G., 2002, Mon. Not. R. Astron. Soc., 333, 730
Coles P., 1989, Mon. Not. R. Astron. Soc., 238, 319

Desjacques V., 2013, Phys. Rev. D, 87, 043505
Gay C., Pichon C., Pogosyan D., 2012, Phys. Rev. D, 85, 023011
Hamaus N., Seljak U., Desjacques V., Smith R. E., Baldauf T., 2010, Phys. Rev. D, 82, 043515
Kac M., 1943, Bull. Am. Math. Soc., 49, 938
Lumsden S. L., Heavens A. F., Peacock J. A., 1989, Mon. Not. R. Astron. Soc., 238, 293
Manera M., Gaztañaga E., 2011, Mon. Not. R. Astron. Soc., 415, 383
McQuinn M., White M., 2015, ArXiv e-prints
Mo H. J., White S. D. M., 1996, Mon. Not. R. Astron. Soc., 282, 347
Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F., Sousbie T., Colombi S., 2009, Mon. Not. R. Astron. Soc., 396, 635
Rice S. O., 1945, Bell System Tech. J., 25, 46
Seljak U., Hamaus N., Desjacques V., 2009, Physical Review Letters, 103, 091303
Sheth R. K., Lemson G., 1999, Mon. Not. R. Astron. Soc., 304, 767
Smith R. E., Desjacques V., Marian L., 2011, Phys. Rev. D, 83, 043526
Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512

PHYSICAL REVIEW D **90,** 103519 (2014)

Statistics of cosmic density profiles from perturbation theory

Francis Bernardeau,[1,2] Christophe Pichon,[2,3,4] and Sandrine Codis[2]

[1]*Institut de Physique Théorique, CEA, IPhT, F-91191 Gif-sur-Yvette, CNRS, URA 2306, F-91191 Gif-sur-Yvette, France*

[2]*Institut d'Astrophysique de Paris and UPMC (UMR 7095), 98 bis boulevard Arago, 75014 Paris, France*

[3]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

[4]*KITP Kohn Hall-4030 University of California, Santa Barbara, California 93106-4030, USA*

(Received 11 November 2013; published 14 November 2014)

The joint probability distribution function (PDF) of the density within multiple concentric spherical cells is considered. It is shown how its cumulant generating function can be obtained at tree order in perturbation theory as the Legendre transform of a function directly built in terms of the initial moments. In the context of the upcoming generation of large-scale structure surveys, it is conjectured that this result correctly models such a function for finite values of the variance. Detailed consequences of this assumption are explored. In particular the corresponding one-cell density probability distribution at finite variance is computed for realistic power spectra, taking into account its scale variation. It is found to be in agreement with Λ-cold dark matter simulations at the few percent level for a wide range of density values and parameters. Related explicit analytic expansions at the low and high density tails are given. The conditional (at fixed density) and marginal probability of the slope—the density difference between adjacent cells—and its fluctuations is also computed from the two-cell joint PDF; it also compares very well to simulations. It is emphasized that this could prove useful when studying the statistical properties of voids as it can serve as a statistical indicator to test gravity models and/or probe key cosmological parameters.

DOI: 10.1103/PhysRevD.90.103519 PACS numbers: 98.80.-k, 98.65.-r

I. INTRODUCTION

With new generations of surveys either from ground-based facilities (e.g. BigBOSS, DES, Pan-STARRS, LSST [1]) or space-based observatories (EUCLID [2], SNAP and JDEM [3]), it will be possible to test with unprecedented accuracy the details of gravitational instabilities, in particular as it enters the nonlinear regime. These confrontations can be used in principle to test the gravity models (see for instance [4,5]) and/or more generally improve upon our knowledge of cosmological parameters as detailed in [2].

There are only a limited range of quantities that can be computed from first principles. Next-to-leading-order terms to power spectra and polyspectra have been investigated extensively over the last few years with the introduction of novel methods. Standard perturbation theory calculations, as described in [6], have indeed been extended by the development of alternative analytical methods that try to improve upon standard calculations. The first significant progress in this line of calculations in the renormalized perturbation theory proposition [7] followed by the closure theory [8] and the time flow equations approach [9]. Latest propositions, namely MPTbreeze [10] and RegPT [11], incorporate 2-loop-order calculations and are accompanied by publicly released codes. Recent developments involve the effective field theory approaches [12].

Alternatively one may look for more global properties of the fields that capture some aspects of their non-Gaussian nature. A number of tests have been put forward from peak statistics (see [13]) that set the stage for Gaussian fields, to topological invariants. The latter, introduced for instance in [14] or in [15], aim at producing robust statistical indicators. This topic was renewed in [16] and [17] with the introduction of the notion of a skeleton. How such observables are affected by weak deviations from Gaussianity was investigated originally in [18] and for instance more recently in [17,19] with the use of standard tools such as the Edgeworth expansion applied here to multiple variable distributions. These approaches, although promising, are hampered by the limited range of applicability of such expansions and as a consequence have to be restricted to a limited range of parameters and are usually confined to the nonrare event region.

There is however at least one counterexample to that general statement: the density probability distribution functions in concentric cells. As we will show in detail in the following it is possible to get a global picture of what the joint density probability distribution function (PDF) should be, including in its rare event tails. The size of the past surveys prevented an effective use of such statistical tools. Their current size makes it now possible to try and confront theoretical calculations with observations.

Hence the aim of the paper is to revisit these calculations and assess their domain of validity with the help of numerical simulations.

To a large extent, the mathematical foundations of the calculation of the density probability distribution functions

in concentric cells are to be found in early works by Balian and Schaeffer [20], who explored the connection between count-in-cell statistics and the properties of the cumulant generating functions. In that paper, the shape of the latter was just assumed without direct connection with the dynamical equations. This connection was established in [21] where it was shown that the leading order generating function of the count-in-cell probability distribution function could be derived from the dynamical equations. More precise calculations were developed in a systematic way in [22] that takes into account filtering effects, as pioneered in [23,24] where the impact of a Gaussian window function or a top-hat window function was taken into account. At the same time, these predictions were subjected to simulations and shown to be in excellent agreement with the numerical results (see for instance [22,25]). We will revisit here the quality of these predictions with the help of more accurate simulations. In parallel, it was shown that the same formalism could address more varied situations: large-scale biasing in [26], projection effects in [27,28]. A comprehensive presentation of these early works can be found in [6].

Insights into the theoretical foundations of this approach were presented in [29] that allow to go beyond the diagrammatic approach that was initially employed. The key argument is that for densities in concentric cells, *the leading contributions* in the implementation of the steepest descent method to the integration over field configurations should be configurations *that are spherically symmetric*. One can then take advantage of Gauss's theorem to map the final field configuration into the initial one with a finite number of initial variables, on a cell-by-cell basis. This is the strategy we adopt below. The purpose of this work is to rederive the fundamental relation that was obtained by the above mentioned authors, and to revisit the practical implementation of these calculations alleviating some of the shortcuts that were used in the literature.

Specifically, the first objective of this paper is to quantity the sensitivity of the predictions for the one-cell PDF for the density on the power spectrum shape, its index and the scale dependence of the latter (the so-called running parameter). The second objective is to show that it is possible to use the two-cell formalism to derive the statistical properties of the density *slope* defined as the difference of the density in two concentric cells of (possibly infinitesimally) close radii and more globally the whole density profile. More specifically we show that for sufficiently steep power spectra (index less than -1), it is possible to take the limit of infinitely close top-hat radii and define the density slope at a given radius. We can then take advantage of this machinery to derive low-order cumulants of this quantity as well as its complete PDF. Finally this investigation allows us to make a theoretical connection with recent efforts (see for instance [30–36]) in exploring the low density regions and their properties [37] such as the constrained average slope and its fluctuations given the (possibly low) value of the local density. This opens the way to exploit the properties of low density regions: we will suggest that the expected profile of low density regions is in fact a robust tool to use when matching theoretical predictions to catalogs.

The outline of the paper is the following. In Sec. II we present the general formalism of how the cumulant generating functions are related to the spherical collapse dynamics. In Sec. III, this relationship is applied to derive the one-point-density PDF; the sensitivity of the predictions with scale and with the power spectrum shape is also reviewed there. In Sec. IV, we define the density profile and the slope, and derive its statistical properties. A summary and discussion on the scope of these results is given in the last section.

II. THE CUMULANT GENERATING FUNCTION AT TREE ORDER

Let us first revisit the derivation of the tree-order cumulant generating functions for densities computed in concentric cells.

A. Definitions and connections to spherical collapse

We consider a cosmological density field, $\rho(\mathbf{x})$, which is statically isotropic and homogeneous. The average value of $\rho(\mathbf{x})$ is set to unity. We then consider a random position $\mathbf{x}_0$ and n concentric cells of radius R_i centered on $\mathbf{x}_0$. The densities, ρ_i, obtained as the density within the radius R_i,

$$\rho_i = \frac{1}{4\pi R_i^3/3}\int_{|\mathbf{x}-\mathbf{x}_0|<R_i} d^3\mathbf{x}\,\rho(\mathbf{x}), \tag{1}$$

form a set of correlated random variables. For a nonlinearly evolved cosmic density field, they display non-Gaussian statistical properties. It is therefore natural to define the generating function of their joint moments as

$$\mathcal{M}(\{\lambda_k\}) = \sum_{p_i=0}^{\infty} \langle \Pi_i \rho_i^{p_i}\rangle \frac{\Pi_i \lambda_i^{p_i}}{\Pi_i p_i!}, \tag{2}$$

which can be simply expressed as

$$\mathcal{M}(\{\lambda_k\}) = \left\langle \exp\left(\sum_i \lambda_i \rho_i\right)\right\rangle. \tag{3}$$

The generating function, $\mathcal{M}(\{\lambda_k\})$, is a function of the n variables λ_k. A very general theorem (see for instance [38,39]) states that this generating function is closely related to the joint *cumulant* generating function,

$$\varphi(\{\lambda_k\}) = \sum_{p_i=0}^{\infty} \langle \Pi_i \rho_i^{p_i}\rangle_c \frac{\Pi_i \lambda_i^{p_i}}{\Pi_i p_i!}, \tag{4}$$

via the relation

$$\mathcal{M}(\{\lambda_k\}) = \exp\left[\varphi(\{\lambda_k\})\right]. \tag{5}$$

Note importantly that this makes $\varphi(\{\lambda_k\})$ an observable on its own [40].

The tree-order expression of such cumulants can be derived from a direct expansion of the density field, i.e.

$$\rho(\mathbf{x}) = 1 + \delta^{(1)} + \delta^{(2)} + \delta^{(3)} + \cdots, \tag{6}$$

where $\delta^{(p)}$ is of order p with respect to the initial density contrast. For Gaussian initial conditions the leading order cumulant (that is the connected parts of the moments) can be derived from the expression of the fields $\delta^{(p)}$. Formally, Wick's theorem imposes [41] that the leading contributions to the p-order cumulant obtained from the following terms:

$$\langle \rho^p \rangle_c = \sum_{\sum_{i=1}^p n_i = 2(p-1)} \langle \Pi_{i=1}^p \delta^{(n_i)} \rangle_c. \tag{7}$$

One of the well-known consequences of that property is that $\langle \rho^p \rangle_c$ scales like $\langle \rho^2 \rangle_c^{2(p-1)}$. It is then natural to define precisely the reduced cumulants, S_p, as

$$S_p(\eta) = \frac{\langle \rho_i^p \rangle_c}{\langle \rho_i^2 \rangle_c^{2(p-1)}}. \tag{8}$$

It has been shown in [21,22] that these quantities are entirely determined by the dynamics of the spherical collapse. More precisely the function ζ that relates the initial density contrast $\tau_{<r}$ within a given shell of radius r to the time-dependent (η) nonlinear density contrast, $\rho_{<R}$ within the shell of radius $R = r\rho_{<R}^{-1/3}$,

$$\rho_{<R} = \zeta(\eta, \tau(<r)), \tag{9}$$

encodes all the necessary ingredients to compute the tree-order cumulants. Note that the mere existence of such a function takes full advantage of Gauss's theorem, as the time evolution of the shell radius depends only on the density contrast at this radius (before shell crossings). More precisely, if one perturbatively expands $\zeta(\eta, \tau)$ with respect to τ,

$$\zeta(\eta, \tau) = \sum_p \nu_p(\eta) \frac{(D_+(\eta)\tau)^p}{p!}, \tag{10}$$

where $D_+(\eta)$ is the linear growth factor between the initial time and time η (with $\nu_0 = 1$, $\nu_1 = 1$), then each $S_p(\eta)$ parameters can be expressed in terms of $\nu_p(\eta)$. For instance

$$S_3(\eta) = 3\nu_2(\eta) + \frac{\mathrm{d}\log\langle\tau^2(r)\rangle}{\mathrm{d}\log r}, \tag{11}$$

$$S_4(\eta) = 4\nu_3(\eta) + 12\nu_2^2(\eta) + (14\nu_2(\eta) - 2)\frac{\mathrm{d}\log[\langle\tau^2(r)\rangle]}{\mathrm{d}\log r} + \frac{7}{3}\left(\frac{\mathrm{d}\log[\langle\tau^2(r)\rangle]}{\mathrm{d}\log r}\right)^2 + \frac{2}{3}\frac{\mathrm{d}^2\log[\langle\tau^2(r)\rangle]}{\mathrm{dlog}^2 r}. \tag{12}$$

The explicit form of $\zeta(\eta, \tau)$ or equivalently the values of $\nu_p(\eta)$ can *a priori* be predicted for any given cosmology. They depend on time—although very weakly—and take simple analytic forms for an Einstein–de Sitter background. For instance, for such a background, we then have $\nu_2 = 34/21$. A more general expression of $\zeta(\eta, \tau)$ can be found in [21,43,44]. In practice one can use a simple expression for $\zeta(\tau)$:

$$\zeta(\tau) = \frac{1}{(1 - D_+(\eta)\tau/\nu)^\nu}. \tag{13}$$

Here we choose $\nu = 21/13 \approx 1.6$ so that the high z skewness of the density contrast is exactly reproduced [45]. We checked that this choice of ν reproduces the exact spherical collapse dynamics for an Einstein–de Sitter background at a precision level of 0.5% from $\zeta = 0.3$ to $\zeta = 2.5$, which is typically the range of values we need to cover.

The understanding of the connection between the leading order statistical properties and the spherical collapse dynamics has been dramatically improved in [26,28,29] where it was realized that it could be extended to the cumulants of any number of concentric cells. We now turn to the presentation of these results.

B. General derivation

We are here interested in the leading order expression of $\varphi = \varphi(\{\lambda_k\})$ for a finite number of concentric cells. In this section we set the dimension of space to be D, having in mind that the formulas we derive should be valid for $D = 2$ or $D = 3$. For completeness, we sketch here the demonstration of the results and refer to [29] for further details. To derive such an expression let us introduce the joint density probability distribution functions, $\mathcal{P}(\{\rho_k\})d\rho_1 \ldots d\rho_n$, so that

$$\exp\left[\varphi(\{\lambda_k\})\right] = \int d\rho_1 \ldots d\rho_n \mathcal{P}(\{\rho_k\}) \exp\left(\sum_i \lambda_i \rho_i\right).$$

This expression can be written in terms of the statistical properties of the *initial* field. Let us define $\tau(\mathbf{x})$ as the initial density contrast. Formally the quantities ρ_i are all functionals of the field, $\tau(\mathbf{x})$ [46], so that the ensemble average of the previous equation can be written as

$$\exp[\varphi] = \int \mathcal{D}\tau(\mathbf{x})\mathcal{P}(\{\tau(\mathbf{x})\}) \exp\left(\sum_i \lambda_i \rho_i(\{\tau(\mathbf{x})\})\right), \tag{14}$$

where we introduced the field distribution function, $\mathcal{P}(\{\tau(\mathbf{x})\})$, and the corresponding measure $\mathcal{D}(\{\tau(\mathbf{x})\})$. These are assumed to be known *a priori*. They depend on the initial conditions and in the following we will assume the initial field is Gaussian distributed [47].

We now turn to the calculation of the generating function at leading order when the overall variance, σ^2, at scale R_i, is small. The idea is to identify the initial field configurations that give the largest contribution to this integral. For convenience, let us assume that the field $\tau(\mathbf{x})$ can be described with a discrete number of variables τ_i. For Gaussian initial conditions, the expression of the joint probability distribution function of τ_i reads

$$\mathcal{P}(\{\tau_k\})d\tau_1 \ldots d\tau_p = \frac{\exp\left[-\Psi(\{\tau_k\})\right]}{\sqrt{(2\pi)^p / \det \Xi}} d\tau_1 \ldots d\tau_p, \tag{15}$$

with

$$\Psi(\{\tau_k\}) = \frac{1}{2}\sum_{ij} \Xi_{ij}\tau_i\tau_j, \tag{16}$$

where Ξ_{ij} is the *inverse* of the covariance matrix, Σ_{ij}, defined as

$$\Sigma_{ij} = \langle \tau_i \tau_j \rangle. \tag{17}$$

The key idea to transform Eq. (14) using Eq. (15) relies on using the steepest descent method. Details of the validity regime of this approach and its construction can be found in [29]. The integral we are interested in is then dominated by a specific field configuration for which the following stationary conditions are verified:

$$\sum_i \lambda_i \frac{\delta \rho_i(\{\tau_k\})}{\delta \tau_j} = \frac{\delta}{\delta \tau_j}\Psi(\{\tau_k\}), \tag{18}$$

for any value of j. Up to this point this is a very general construction. Let us now propose a solution to these stationary equations that is consistent with the class of spherically symmetric problems we are interested in. The main point is the following: the configurations that are solutions of this equation, that is the values of $\{\tau_k\}$, depend specifically on the choice of the functionals $\rho_i(\{\tau_k\})$. When these functionals correspond to spherically symmetric quantities, the corresponding configurations are also likely to be spherically symmetric. But then Gauss's theorem is making things extremely simple: *before shell crossing*, each of the final density ρ_i can indeed be expressed in terms of a *single* initial quantity, namely the linear density contrast of the cell centered on $\mathbf{x}_0$ that contained the same amount of matter in the initial density field. We denote τ_i the corresponding density contrast, which means that, following definition (9), we have

$$\rho_i = \zeta(\eta, \tau_i), \tag{19}$$

and τ_i is the amplitude of the initial density within a specific radius [49], r_i, which obeys $r_i = R_i\rho_i^{1/D}$ thanks to mass conservation. The specificity of this mapping implies in particular that

$$\frac{\delta \rho_i(\{\tau_k\})}{\delta \tau_j} = \delta_{ij}\zeta'(\tau_i), \tag{20}$$

so that the stationary conditions (18) now read

$$\lambda_j \zeta'(\tau_j) = \frac{\delta}{\delta \tau_j}\Psi(\{\tau_k\}). \tag{21}$$

Note that the no-shell crossing conditions imply that if $R_i < R_j$, then $r_i < r_j$, which in turn implies that

$$\rho_i < \rho_j (R_j/R_i)^D. \tag{22}$$

It follows that the parameter space $\{\rho_k\}$ is not fully accessible. In the specific example we explore in the following, this restriction is not significant, but it could be in some other cases.

We are now close to the requested expression for $\varphi(\{\lambda_k\})$ as we have

$$\exp\left[\varphi(\{\lambda_k\})\right] = \int \mathrm{d}\tau_1 \ldots \mathrm{d}\tau_n \mathcal{P}(\{\tau_k\}) \exp\left(\sum_i \lambda_i \rho_i(\{\tau_k\})\right).$$

To get the leading order expression of this form for $\varphi(\{\lambda_k\})$, using the steepest descent method, one is simply requested to identify the quantities that are exponentiated. As a result we have

$$\varphi(\{\lambda_k\}) = \sum_i \lambda_i \rho_i - \Psi(\{\rho_k\}), \tag{23}$$

where ρ_i are determined by the stationary conditions (21). The latter can be written equivalently as

$$\lambda_i = \frac{\partial}{\partial \rho_i}\Psi(\{\rho_k\}), \tag{24}$$

when all quantities are expressed in terms of ρ_i. Equation (24) is the general expression that we will exploit in the following. Formally, note that (21)–(23) imply that $\varphi(\{\lambda_k\})$ is the Legendre transform of Ψ when the latter is seen as a function of ρ_i, that is

$$\Psi(\{\rho_k\}) = \frac{1}{2}\sum_{ij}\Xi_{ij}(\{\rho_k\})\tau(\rho_i)\tau(\rho_j), \tag{25}$$

where the functional form $\tau(\rho)$ is obtained from the inversion of (19) at a fixed time, and Ξ_{ij} is the inverse matrix of the cross-correlation of the density in cells of radius $R_i\rho_i^{1/D}$ [cf. Eq. (17)]:

$$\Sigma_{ij} = \left\langle \tau\left(< R_i\rho_i^{1/D}\right)\tau\left(< R_j\rho_j^{1/D}\right)\right\rangle, \tag{26}$$

$$\sum_j \Sigma_{ij}\Xi_{jk} = \delta_{ik}. \tag{27}$$

These coefficients therefore depend on the whole set of both radii R_i and densities ρ_i. From the properties of Legendre transform, it follows in particular that

$$\rho_i = \frac{\partial}{\partial\lambda_i}\varphi(\{\lambda_k\}). \tag{28}$$

Although known for more than a decade, Eqs. (23)–(28) and their consequences have not been exploited to their full power in the literature. This is partially what we intend to do in this paper and in subsequent ones. For now, in order to get better acquainted with this formalism, let us first explore some of its properties.

C. General formalism

The relation Eqs. (23)–(28) have been derived for Gaussian initial conditions. This eases the presentation but it is not a key assumption. For instance in Eq. (15), $\Psi(\{\tau_k\})$ does not need to be quadratic in τ_k as for Gaussian initial conditions. If the initial conditions were to be non-Gaussian these features would have to be incorporated in the expression of $\Psi(\{\tau_k\})$. It would not however change the functional relation between $\varphi(\{\lambda_k\})$ and $\Psi(\{\tau_k\})$, provided $\Psi(\{\tau_k\})$ is properly defined when the variance is taken in its zero limit.

One can then observe that the Legendre transform between these two functions can be inverted [50]. Applying the fundamental relation at precisely the initial time, in a regime where $\rho_i \approx 1 + D_+(\eta)\tau_i$, will give the expression of the function $\Psi(\{\tau_k\})$ in terms of the initial cumulant generating function.

One can actually pursue this idea more generally. Let us define the nonlinear spherical transform $\zeta_\rho(\eta;\rho_0,\eta_0)$ that gives the value of the density ρ within a given radius R at time η knowing the density ρ_0 at time η_0 within radius $R_0 = R(\rho/\rho_0)^{1/3}$. It is obtained after τ has been eliminated in

$$\zeta_\rho(\eta;\rho_0,\eta_0) = \zeta(\eta,\tau), \tag{29}$$

$$\rho_0 = \zeta(\eta_0,\tau), \tag{30}$$

where ζ is defined in Eq. (9). Using the form (13), one gets

$$\zeta(\eta;\rho_0,\eta_0)^{-1/\nu} - 1 = \frac{D_+(\eta)}{D_+(\eta_0)}(\rho_0^{-1/\nu} - 1). \tag{31}$$

Incidentally we can note that the inverse function is obtained by changing η into η_0.

Then the general formulation of our result is that the Legendre transform of the joint cumulant generating function for a choice of radii R_k and taken at time η, which we denote here as $\Psi(\{(\rho_k,R_k)\},\eta)$, can be expressed in terms of the same Legendre transform taken at any other time, η',

$$\Psi(\{(\rho_k,R_k)\},\eta) = \Psi\left(\left\{\zeta_\rho(\eta';\rho_k,\eta), R_k\frac{\rho_k^{1/3}}{\zeta_\rho^{1/3}(\eta';\rho_k,\eta)}\right\},\eta'\right). \tag{32}$$

This is a general formalism that encompasses the result we just described [51] but can also be applied for any initial conditions or any time as it does refer explicitly to the initial conditions.

In this paper we will however use this construction for initial Gaussian conditions only with explicit use of the expressions derived in the previous subsection.

D. Scaling relations

It is interesting to note that the cumulant generating function has a simple dependence on the overall amplitude of the correlators σ_0^2. Let us denote in this subsection $\varphi_{\sigma_0}(\{\lambda_k\})$ the value of the cumulant generating function for a fixed value of σ_0. It is then straightforward to express $\varphi_{\sigma_0}(\{\lambda_k\})$ in terms of $\varphi_1(\{\lambda_k\})$, the expression of the generating function when σ_0 is set to unity. Indeed $\Psi(\{\rho_k\})$ is inversely proportional to σ_0^2 for fixed values of ρ_k. As a result λ_k scale like $1/\sigma_0^2$ for fixed values of $\{\rho_k\}$. Note that we have the following identity:

$$\varphi_{\sigma_0}(\{\lambda_k\}) = \frac{1}{\sigma_0^2}\varphi_1(\{\lambda_k/\sigma_0^2\}), \tag{33}$$

while the variables ρ_k are independent of σ_0.

In the upcoming applications we will make use of this property as we will keep the overall normalization as a free parameter—that will eventually be adjusted on numerical results, but will use the structural form of $\varphi_1(\{\lambda_k\})$ as predicted from the general theory. In particular this structural form depends on the specific shape of the power spectrum through the cross-correlation matrix Σ_{ij}.

E. The one-cell generating function

Turning back to the application of Eqs. (23)–(28), one obvious simple application corresponds to the one-cell characteristic function. In this case

$$\Psi(\rho) \equiv \frac{1}{2\sigma^2(R\rho^{1/D})}\tau(\rho)^2, \tag{34}$$

where

$$\sigma^2(r) = \langle \tau(<r)\tau(<r)\rangle. \tag{35}$$

The Legendre transform is then straightforward and $\varphi(\lambda)$ takes the form

$$\varphi(\lambda) = \lambda\rho - \frac{1}{2\sigma^2(R\rho^{1/D})}\tau(\rho)^2, \tag{36}$$

with ρ computed implicitly as a function of λ via Eq. (24). One way of rewriting this equation is to define $\tau_{\rm eff} = \tau\sigma(R)/\sigma(R\rho^{1/3})$ and the function $\zeta_{\rm eff}(\tau_{\rm eff})$ through the implicit form,

$$\zeta_{\rm eff}(\tau_{\rm eff}) = \zeta(\tau) = \zeta\left(\tau_{\rm eff}\frac{\sigma(R\zeta_{\rm eff}^{1/D})}{\sigma(R)}\right). \tag{37}$$

Then the expression of $\varphi(\lambda)$ is given by

$$\varphi(\lambda) = \lambda\rho - \frac{1}{2\sigma^2(R)}\tau_{\rm eff}^2, \tag{38}$$

with the stationary condition

$$\tau_{\rm eff} = \lambda\sigma^2\zeta'_{\rm eff}(\hat{\tau}). \tag{39}$$

In [22], the expression of the cumulant generating function was presented with this form. This is also the functional form one gets when one neglects the filtering effects (as was initially done in [21]) or for the so-called nonlinear hierarchical model used in [52]. Note that it is not possible however to use such a remapping for more than one cell. Note finally that this is a precious formulation for practical implementations, as one may rely on fitted forms for $\zeta_{\rm eff}$ to construct the generating function $\varphi(\lambda)$ while preserving its analytical properties. It is indeed always possible, once one has been able to numerically compute $\varphi(\lambda)$ for specific values of λ, to define $\zeta_{\rm eff}$ by Legendre transform and construct a fitted form with low-order polynomials while this is not possible for $\varphi(\lambda)$ which exhibits nontrivial analytical properties as we will see later on. This approach was used in [28]. It is also this procedure we use in Sec. IV for constructing the profile PDF.

F. Recovering the PDF via inverse Laplace transform

In the following we will exploit the expression for the cumulant generating function to get the one-point and joint density PDFs. To avoid confusion with the variables ρ_i that appear in the expression of Ψ, we will use the superscript $\widehat{\ }$ to denote measurable densities, the PDF of which we wish to compute.

In general, the joint density PDF, $\mathcal{P} = \mathcal{P}(\hat{\rho}_1, \ldots, \hat{\rho}_n)$, that gives the probability that the densities within a set of n concentric cells of radii $R_1, \ldots, R_n$ are $\hat{\rho}_1, \ldots, \hat{\rho}_n$ within $d\hat{\rho}_1 \ldots d\hat{\rho}_n$ is given by

$$\mathcal{P} = \int_{-i\infty}^{+i\infty}\frac{{\rm d}\lambda_1}{2\pi i}\cdots\frac{{\rm d}\lambda_n}{2\pi i}\exp\left(-\sum_i \lambda_i\hat{\rho}_i + \varphi(\{\lambda_k\})\right). \tag{40}$$

where the integration in λ_i should be performed in the complex plane so as to maximize convergence. This equation defines the inverse Laplace transform of the cumulant generating function [53]. In the one-cell case, Eq. (40) simply reads

$$\mathcal{P}(\hat{\rho}_1) = \int_{-i\infty}^{+i\infty}\frac{{\rm d}\lambda_1}{2\pi i}\exp(-\lambda_1\hat{\rho}_1 + \varphi(\lambda_1)), \tag{41}$$

i.e. the PDF is the inverse Laplace transform of the one-variable moment generating function. This inversion is known to be tricky, and to our knowledge there are no known general foolproof methods. One practical difficulty is that it generically relies on the analytic continuation of the predicted cumulant generating function in the complex plane. It is therefore crucial to have a good knowledge of the analytic properties of $\varphi(\lambda)$, which is typically difficult since $\varphi(\lambda)$ is defined itself as the Legendre transform of $\Psi(\rho)$. Only a limited set of $\Psi(\rho)$ yield analytical $\varphi(\lambda)$, which in turn can be inverse Laplace transformed.

III. THE ONE-POINT PDF

Up to this point, the whole construction presented in the previous section would be a mere mathematical trick to compute explicit cumulants for top-hat window functions sparing the pain of lengthy integrations on wave modes. In this paper, we furthermore aim to use the cumulant generating function computed in the uniform limit $\Sigma_{ij} \to 0$ as an approximate form for the *exact* generating function when the Σ_{ij} are finite (but small). Note that this is a nontrivial extension for which we have no precise mathematical justifications. It assumes that the global properties of $\varphi(\{\lambda_k\})$—and in particular its analytical properties (which will be of crucial importance in the following)—should be meaningful for *finite values of* λ_k, and not only in the vicinity of $\{\lambda_k = 0\}$.

We now conjecture without further proof that they correctly represent the cumulant generating function for finite values of the variance.

A. General formulas and asymptotic forms

The implementation of the quadrature in Eq. (41) has been attempted in various papers [21,28,52], relying on different hypotheses for $\varphi(\lambda)$ [54]. Figure 1 yields a graphical representation of the stationary equation for a power law model with index $n=-1.5$. The implicit equation, $\Psi'[\rho]=\lambda$, always has a solution in the vicinity of $\rho\approx 0$. Expanding this equation around this point naturally gives the low-order cumulants at an arbitrary order.

Figure 1 shows graphically that there is a maximum value for λ, λ_c that can be reached, so that the Legendre transform of Ψ is not defined for $\lambda>\lambda_c$. It corresponds to a value $\rho=\rho_c$. At this location we have

$$0=\Psi''[\rho_c],\qquad \lambda_c=\Psi'[\rho_c]. \tag{42}$$

Note that at $\rho=\rho_c$, Ψ is regular [in particular, the corresponding singular behavior in $\varphi(\lambda)$ is not related to any singularity of the spherical collapse dynamics]. The function $\varphi(\lambda)$ can be expanded at this point. In other words, Eq. (24) can be inverted as a series near (ρ_c,λ_c) [where Eq. (42) holds], and integrated for $\varphi(\lambda)$ using Eq. (28). We give here a whole set of subleading terms that we will take advantage of in the following,

$$\begin{aligned}\varphi(\lambda)=\varphi_c+(\lambda-\lambda_c)\rho_c+\frac{2}{3}\sqrt{\frac{2}{\pi_3}}(\lambda-\lambda_c)^{3/2}-\frac{\pi_4(\lambda-\lambda_c)^2}{6\pi_3^2}\\ +\frac{(\frac{1}{\pi_3})^{7/2}(5\pi_4^2-3\pi_3\pi_5)(\lambda-\lambda_c)^{5/2}}{45\sqrt{2}}-\frac{(40\pi_4^3-45\pi_3\pi_5\pi_4+9\pi_3^2\pi_6)(\lambda-\lambda_c)^3}{810\pi_3^5}\\ +\frac{(\frac{1}{\pi_3})^{13/2}(385\pi_4^4-630\pi_3\pi_5\pi_4^2+168\pi_3^2\pi_6\pi_4+3\pi_3^2(35\pi_5^2-8\pi_3\pi_7))(\lambda-\lambda_c)^{7/2}}{7560\sqrt{2}}+\cdots,\end{aligned} \tag{43}$$

where $\pi_n=\partial^n\Psi/\partial\rho^n(\rho_c)$. It is to be noted that the leading singular term scales like $(\lambda-\lambda_c)^{3/2}$. The coefficients π_i are all related to the function Ψ and are therefore (cosmological) model dependent [55].

What are the consequences of this behavior for the PDF of the density? Let us present analytical forms for the inverse Laplace transform of $\exp\varphi$. The idea is that the inverse transform can be obtained via a saddle point approximation of Eq. (41) assuming the variance is small. Formally it leads to the conditions that should be met at the saddle point λ_s [56],

$$\frac{\partial}{\partial\lambda}[\lambda\hat{\rho}-\varphi(\lambda)]=0, \tag{44}$$

$$\frac{\partial^2}{\partial\lambda^2}[\lambda\hat{\rho}-\varphi(\lambda)]<0. \tag{45}$$

The first condition leads to $\rho(\lambda_s)=\hat{\rho}$, the second to $\lambda_s<\lambda_c$. This condition simply means that this approximation can be used if $\hat{\rho}<\rho_c$. The resulting simple expression for the density PDF is

$$\mathcal{P}(\hat{\rho})=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{\partial^2\Psi(\hat{\rho})}{\partial\hat{\rho}^2}}\exp\left[-\Psi(\hat{\rho})\right]. \tag{46}$$

It is valid as long as the expression that appears in the square root is positive, i.e. $\hat{\rho}<\rho_c$. When this condition is not satisfied, the singular behavior of φ near λ_c dominates the integral in the complex plane. This leads to the following expression for $\mathcal{P}(\hat{\rho})$ as described in Appendix B 2,

$$\begin{aligned}\mathcal{P}(\hat{\rho})\approx\exp\left(\varphi_c-\lambda_c\hat{\rho}\right)\left(\frac{3\Im(a_{\frac{3}{2}})}{4\sqrt{\pi}(\hat{\rho}-\rho_c)^{5/2}}+\frac{15\Im(a_{\frac{5}{2}})}{8\sqrt{\pi}(\hat{\rho}-\rho_c)^{7/2}}\right.\\ \left.+\frac{105(\Im(a_{\frac{3}{2}})a_2+\Im(a_{\frac{7}{2}}))}{16\sqrt{\pi}(\hat{\rho}-\rho_c)^{9/2}}+\cdots\right),\end{aligned} \tag{47}$$

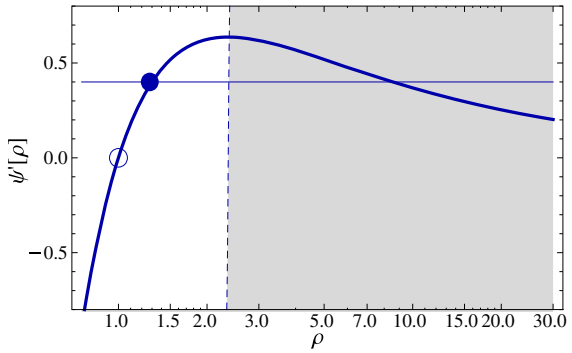


FIG. 1 (color online). A graphical representation of the one-dimensional stationary condition $\lambda=\Psi'[\rho]$. There is a maximum value for λ that corresponds to a critical value ρ_c for ρ defined in Eq. (42).

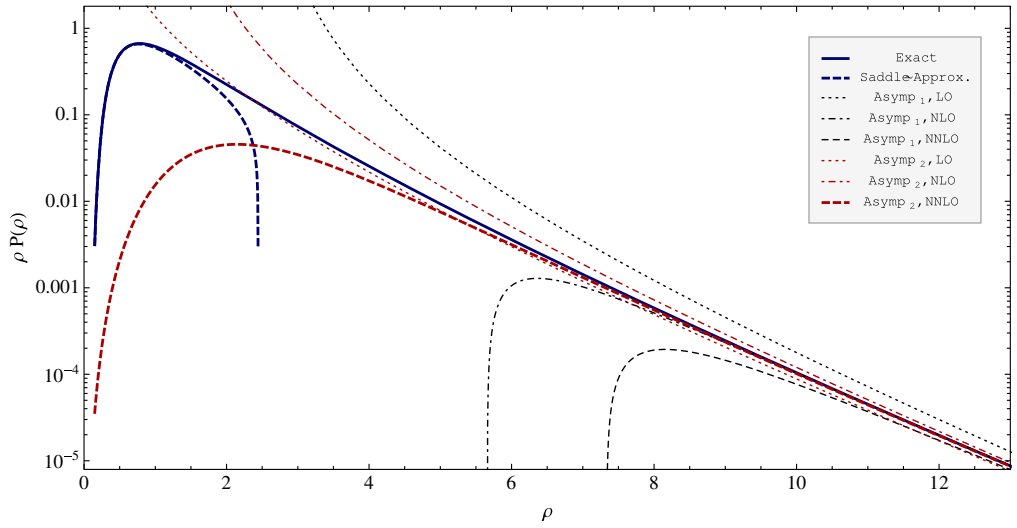


FIG. 2 (color online). The PDF of the one-point density. The blue solid line is the numerical integration, the red dashed line the low ρ asymptotic form of Eq. (46); the other lines correspond to the large ρ asymptotic forms proposed in the text: the dark lines correspond to the form (47) and the red lines to the form (48) and the forms are computed at leading order, at next-to-leading, and next-to-next-to-leading order for respectively the dotted, dot-dashed and dashed curves. The plots are given for $\sigma^2 = 0.45$ and a power law index of $n = -1.5$.

where a_j are the coefficients in front of $(\lambda - \lambda_c)^j$ in Eq. (43), (e.g. $a_{3/2} = 2/3\sqrt{2/\pi_3}$) and $\Im()$ is the imaginary part. Equation (47) has an exponential cutoff at large $\hat{\rho}$ scaling like $\exp(\lambda_c\hat{\rho})$. This property is actually robust and is preserved when one performs the inverse Laplace transform for finite values of the variance, or even large value of the variance (see [20,57]). It also gives a direct transcription of why $\varphi(\lambda)$ becomes singular: for values of λ that are larger than λ_c, the integral $\int \mathrm{d}\hat{\rho}\mathcal{P}(\hat{\rho})\exp(\lambda\hat{\rho})$ is not converging.

Note that in practice, it is best to rely on an alternative asymptotic form to Eq. (47) that is better behaved and remains finite for $\hat{\rho} \to \rho_c$. It is built in such a way that it has the same asymptotic behavior as Eq. (47) at a given order in the large ρ limit. The following form,

$$\mathcal{P}(\hat{\rho}) = \frac{3a_{\frac{3}{2}}\exp(\varphi_c - \lambda_c\hat{\rho})}{4\sqrt{\pi}(\hat{\rho} + r_1 + r_2/\hat{\rho} + \cdots)^{5/2}}, \tag{48}$$

where the r_i parameters are adjusted to fit the results of the previous expansion, proved very robust. At next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) we have

$$r_1 = -\frac{\Im(a_{\frac{5}{2}})}{\Im(a_{\frac{3}{2}})} - \rho_c, \tag{49}$$

$$r_2 = -\frac{7(2a_2a_{\frac{3}{2}}^2 + 2a_{\frac{7}{2}}a_{\frac{3}{2}} - a_{\frac{5}{2}}^2)}{4a_{\frac{3}{2}}^2}. \tag{50}$$

However, none of these asymptotic forms are accurate for the full range of density values; in general one has to rely on numerical integrations in the complex plane which can be done accurately and quickly, as described in Appendix B. The comparison between the analytical forms and the numerical integrations are shown in Fig. 2. Such comparisons are in fact conversely useful to assess the precision of the numerical integrations. Note that for the case explicitly shown, which corresponds to $\sigma^2 = 0.45$ and a power law index of $n = -1.5$, the asymptotic forms (46) and (48) at NNLO are valid within 2% everywhere but for the range $1 < \rho < 10$, where one must rely on an explicit integration in the complex plane.

B. Practical implementation, comparisons with *N*-body results

We now move to an explicit comparison of these predictions to N-body results. The simulations are described in Appendix D. They are determined in particular by the linear power spectrum $P^{\mathrm{lin}}(k)$ set for the initial conditions. The knowledge of the power spectrum determines the values of the cross-correlation matrix, $\Sigma_{ij}(R_i, R_j)$, that are explicitly given by

$$\Sigma_{ij} = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} P^{\mathrm{lin}}(k) W_{\mathrm{3D}}(kR_i) W_{\mathrm{3D}}(kR_j), \tag{51}$$

where $W_{\mathrm{3D}}(k)$ is the shape of the top-hat window function in Fourier space,

$$W_{\mathrm{3D}}(k) = 3\sqrt{\frac{\pi}{2}}\frac{J_{3/2}(k)}{k^{3/2}}, \tag{52}$$

where $J_{3/2}(k)$ is the Bessel function of the first kind of index 3/2. In three dimensions (3D), it is actually possible to express $W_{\mathrm{3D}}(k)$ in terms of elementary functions as

$$W_{3D}(k) = \frac{3}{k^2}(\sin(k)/k - \cos(k)). \tag{53}$$

For the one-cell case we only need to know the amplitude and scale dependence of σ_R^2 defined as

$$\sigma^2(R) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} P^{\rm lin}(k) W_{3D}^2(kR). \tag{54}$$

To a first approximation, $\sigma^2(R)$ can be parametrized with a simple power law $\sigma^2(R) \sim R^{-(n_s+3)}$. It is this functional form which was used in the previous section. The detailed predictions of the PDF depend however on the precise scale dependence of $\sigma^2(R)$. Such scale dependence can be computed numerically from the shape of the power spectrum but then makes it difficult to derive the function $\varphi(\lambda)$ from the Legendre transform. So in order to retain simple analytic expressions for the whole cumulant generating function, we adopt a simple prescription for the scale dependence of $\sigma^2(R)$ given by

$$\sigma^2(R) = \frac{2\sigma^2(R_p)}{(R/R_p)^{n_1+3} + (R/R_p)^{n_2+3}}, \tag{55}$$

where R_p is a pivot scale. Such a parametrization ensures that the single-point $\Psi(\rho)$ function takes a simple *analytic* form as it involves the inverse of $\sigma^2(R)$. Note that our ansatz can be extended to an arbitrary (finite) number of terms in the denominator.

The values of the three parameters, $\sigma^2(R_p)$, n_1 and n_2 are then adjusted so that the model reproduces (i) the *measured* variance $\sigma^2(R)$, (ii) the linear theory index

$$n(R) = -3 - \frac{\mathrm{d}\log(\sigma(R))}{\mathrm{d}\log R}, \tag{56}$$

and (iii) its running parameter

$$\alpha(R) = \frac{\mathrm{d}\log(n(R))}{\mathrm{d}\log R}, \tag{57}$$

at the chosen filtering scale. It is important to point out that we do not take the amplitude of $\sigma^2(R)$ as predicted by linear theory. We consider instead its overall amplitude as a free parameter and $\sigma^2(R)$ is directly measured from the N-body results. The reason is that using the predicted value of $\sigma^2(R)$ would simply introduce too large errors and this dependence can always be scaled out using the relation of Sec. II D [58].

In Fig. 3, we explicitly show the comparison between our predictions following the prescription we just described to measured PDFs. The predictions show a remarkable agreement with the measured PDF. Recall that only one parameter, σ_R, is adjusted to the numerical data. In particular the predictions reproduce with an extremely good accuracy the PDF tails in both the low density and high density regions. The plot of the residuals shows the predictions are at the percent level over a large range of density values. And this result is obtained for a squared variance close to 0.5.

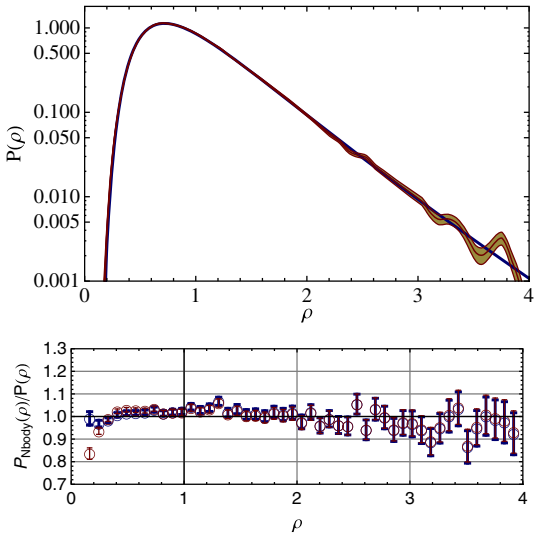


FIG. 3 (color online). Comparison with simulations (top panel) with residuals (bottom panel). The solid line is the theoretical prediction computed for a variance of $\sigma_R^2 = 0.47$ as measured in the simulation, a power law index of $n = -1.576$ and a running parameter $\alpha = 0.439$ corresponding to the input linear power spectrum. The measured PDF in the simulation is shown as a band corresponding to its 1–σ error bar (but different data points are correlated). The residuals show the ratio of the measured PDF in bins with the predictions (computed in bins as well). The thin red symbols show the comparison when the running parameter is set to zero in the prediction.

More extended comparisons with numerical simulations are shown on Fig. 15 which qualifies in more detail the validity regime of our predictions. Note that up to $\sigma = 0.64$ ($\sigma^2 = 0.41$), we see no significant departure from the results of the simulation in the whole range of available densities, that is in particular up to about the 5σ rare event in the high density tail. This success is to be contrasted with the Edgeworth expansion approach which breaks for $|\delta| \geq \sigma$ (see for instance [59]).

We observe that departures from our calculations start to be significant, of the order of 10%, when $\sigma^2(R)$ is of the order of 0.7 or more [60]. These results also show that taking into account the scale dependence of the local index through the introduction of the running parameter improves upon the predictions in the low density region. This aspect is examined in more detail in Fig. 4 which shows the ratio of the predicted PDFs with and without taking into account the running parameter. We see that the PDFs are mostly affected on their tails. This is related to the fact that the kurtosis is the lowest order cumulant to be changed when one introduces a running parameter [24], as can be verified

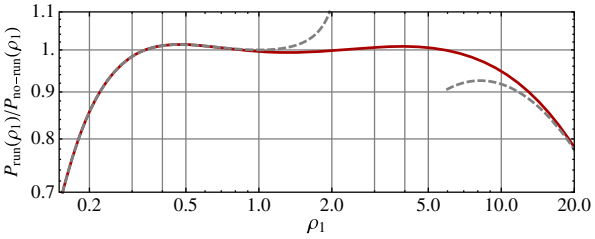


FIG. 4 (color online). Ratio of the one-point density PDF when the running parameter is taken into account over the PDF when it is not. The running model is the same as in the previous plot. The dashed lines are the ratio of the corresponding asymptotic forms in the low and high density regions.

from the relation (12). The effect is actually detectable in the low density region only and confirms the fact that the introduction of a running parameter can have a noticeable impact when comparisons at the percent level are to be done.

IV. THE STATISTICAL PROPERTIES OF THE DENSITY SLOPE AND PROFILE

We now move to the application of the general formalism to the two-cell case. Such situations have already been encountered in [26] to compute effective bias properties, and in [28] to compute the aperture mass statistics out of two concentric angular cells of fixed radius ratio. But all these applications eventually reduce to an effective one-cell case. We are interested here in genuine two-cell statistics.

Let us first make a remark that may seem trivial. Indeed, from the very definition of cumulant generating functions, one should have

$$\varphi_{2-\rm cell}(\lambda_1,\lambda_2=0)=\varphi_{1-\rm cell}(\lambda_1), \tag{58}$$

where $\varphi_{2-\rm cell}(\lambda_1,\lambda_2)$ is the cumulant generating function for cells of radii R_1 and R_2 and $\varphi_{1-\rm cell}(\lambda_1)$ is the cumulant generating function for one cell of radius R_1. Checking that the relations (23)–(28) verify this property makes a sound mathematical exercise. More generally one can show that our formulation is consistent with radii decimation, that is when one computes the cumulant generating functions of a restricted number of variables out of a larger number one gets a consistent result. The demonstration of this property is given in Appendix A.

The purpose of this section is now to define the statistical properties of the density profile, while relying on the fact that the function $\varphi(\lambda_1,\lambda_2)$ has a well-defined, but nontrivial limit, when one sets $\Delta R=R_2-R_1\ll R_1$.

A. The density slope

From the densities in two concentric cells, it is indeed always possible to define the corresponding density slope as

$$\hat{s}(R_1,R_2)=\frac{R_1}{\Delta R}[\hat{\rho}_2-\hat{\rho}_1]. \tag{59}$$

In the limit of a vanishing smoothing radius difference, $\hat{s}$ will define the local density slope. In the following we will in particular see that this is a genuine limit in the sense that it leads to regular and nontrivial expressions.

Let us start with basic preliminary calculations; to avoid too complicated notations, let us define

$$\sigma^2_{R_1}\equiv\sigma^2(R_1,R_1), \tag{60}$$

$$\sigma^2_{R_1R_2}\equiv\sigma^2(R_1,R_2), \tag{61}$$

$$\sigma^2_{R_2}\equiv\sigma^2(R_2,R_2), \tag{62}$$

which are quantities involved in the expressions of cumulants. The variance of $\hat{s}$ is then for instance given by

$$\langle\hat{s}^2\rangle=\left(\frac{R_1}{\Delta R}\right)^2(\sigma^2_{R_1}-2\sigma^2_{R_1R_2}+\sigma^2_{R_2}). \tag{63}$$

From the general theory, Eqs. (23)–(28) implemented for two cells, one can compute the generating function of joint density contrasts in concentric cells [61] in the limit of small λ_i. Up to third order it is explicitly given by

$$\begin{aligned}\varphi(\lambda_1,\lambda_2)=&\lambda_1+\lambda_2+\frac{1}{2}\lambda_1^2\sigma^2_{R_1}+\frac{1}{2}\lambda_2^2\sigma^2_{R_2}+\lambda_1\lambda_2\sigma^2_{R_1R_2}+\lambda_1^3\left(\frac{1}{2}\nu_2\sigma^4_{R_1}+\frac{1}{6}R_1\sigma^2_{R_1}\frac{\mathrm{d}}{\mathrm{d}R_1}\sigma^2_{R_1}\right)+\lambda_2^3\left(\frac{1}{2}\nu_2\sigma^4_{R_2}+\frac{1}{6}R_2\sigma^2_{R_2}\frac{\mathrm{d}}{\mathrm{d}R_2}\sigma^2_{R_2}\right)\\&+\lambda_1^2\lambda_2\left(\frac{1}{2}\nu_2\sigma^2_{R_1R_2}(\sigma^2_{R_1,R_2}+2\sigma^2_{R_1})+\frac{1}{6}\left(2R_1\sigma^2_{R_1}\frac{\partial}{\partial R_1}\sigma^2_{R_1R_2}+\sigma^2_{R_1R_2}\left(2R_2\frac{\partial}{\partial R_2}\sigma^2_{R_1R_2}+R_1\frac{\mathrm{d}}{\mathrm{d}R_1}\sigma^2_{R_1}\right)\right)\right)\\&+\lambda_1\lambda_2^2\left(\frac{1}{2}\nu_2\sigma^2_{R_1R_2}(\sigma^2_{R_1R_2}+2\sigma^2_{R_2})+\frac{1}{6}\left(2R_2\sigma^2_{R_2}\frac{\partial}{\partial R_2}\sigma^2_{R_1R_2}+\sigma^2_{R_1R_2}\left\{2R_1\frac{\partial}{\partial R_1}\sigma^2_{R_1R_2}+R_2\frac{\mathrm{d}}{\mathrm{d}R_2}\sigma^2_{R_2}\right\}\right)\right),\end{aligned} \tag{64}$$

where $\nu_2=34/21$ for a 3D dynamics in an Einstein–de Sitter background. In Eq. (64), the cumulants and joint cumulants can be read out using definition (4) or via differentiation. For instance,

$$\langle\hat{\rho}_1^3\rangle_c = 3\nu_2\sigma_{R_1}^4 + \sigma_{R_1}^2\frac{R_1\mathrm{d}}{\mathrm{d}R_1}\sigma_{R_1}^2, \tag{65}$$

$$\langle\hat{\rho}_1^2\hat{\rho}_2\rangle_c = \nu_2\sigma_{R_1R_2}^2(\sigma_{R_1R_2}^2 + 2\sigma_{R_1}^2) + \frac{2}{3}\sigma_{R_1}^2\frac{R_1\partial}{\partial R_1}\sigma_{R_1R_2}^2 + \frac{1}{3}\sigma_{R_1R_2}^2\left(2\frac{R_2\partial}{\partial R_2}\sigma_{R_1R_2}^2 + \frac{R_1\mathrm{d}}{\mathrm{d}R_1}\sigma_{R_1}^2\right), \tag{66}$$

and the cumulants $\langle\rho_1\rho_2^2\rangle_c$ and $\langle\rho_2^3\rangle_c$ can be obtained exchanging the role of R_1 and R_2. It is then also possible to derive the explicit form for a number of auto- and cross-cumulants between the density $\hat{\rho} \equiv \hat{\rho}_1$ in the first cell and the slope $\hat{s}$ as defined in (59). For instance,

$$\langle\hat{\rho}^2\hat{s}\rangle_c = \frac{R_1}{\Delta R}[\langle\hat{\rho}_1^2\hat{\rho}_2\rangle_c - \langle\hat{\rho}_1^3\rangle_c], \tag{67}$$

$$\langle\hat{\rho}\hat{s}^2\rangle_c = \left[\frac{R_1}{\Delta R}\right]^2[\langle\hat{\rho}_1\hat{\rho}_2^2\rangle_c - 2\langle\hat{\rho}_1^2\hat{\rho}_2\rangle_c + \langle\hat{\rho}_1^3\rangle_c], \tag{68}$$

$$\langle\hat{s}^3\rangle_c = \left[\frac{R_1}{\Delta R}\right]^3[\langle\hat{\rho}_2^3\rangle_c - 3\langle\hat{\rho}_1\hat{\rho}_2^2\rangle_c + 3\langle\hat{\rho}_1^2\hat{\rho}_2\rangle_c - \langle\hat{\rho}_1^3\rangle_c]. \tag{69}$$

Following the one-cell case (see for instance [6]) it is possible to formally define the reduced cross-correlations that are independent on the overall amplitude of the power spectrum. More precisely, the reduced cross-correlations can be defined as

$$S_{p0} = \frac{\langle\hat{\rho}^p\rangle_c}{\langle\hat{\rho}^2\rangle_c^{p-1}}, \tag{70}$$

$$S_{pq} = \frac{\langle\hat{\rho}^p\hat{s}^q\rangle_c}{\langle\hat{\rho}^2\rangle_c^{p-1}\langle\hat{\rho}\,\hat{s}\rangle_c\langle\hat{s}^2\rangle_c^{q-1}}, \tag{71}$$

$$S_{0q} = \frac{\langle\hat{s}^q\rangle_c}{\langle\hat{s}^2\rangle_c^{q-1}}. \tag{72}$$

From the previous expressions these quantities can be computed in the limit of an infinitely small variance.

B. Cumulants and slope in the limit $(\Delta R)/R \to 0$

Let us now consider the statistical properties of $\hat{s}$ in the limit $(\Delta R)/R \to 0$. To start with, let us compute the variance of the slope $\hat{s}$ in the limit $\Delta R/R \to 0$. Its variance is formally given by

$$\langle\hat{s}^2\rangle = \frac{R_1^2\partial^2}{\partial R_1\partial R_2}\sigma_{R_1R_2}^2\Big|_{R=R_1=R_2}. \tag{73}$$

This expression can easily be expressed in terms of the power spectrum,

$$\langle\hat{s}^2\rangle = \int\frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}P^{\mathrm{lin}}(k)\tilde{W}_{3\mathrm{D}}^2(kR), \tag{74}$$

where $\tilde{W}_{3\mathrm{D}}(k)$ is the logarithmic derivative of $W_{3\mathrm{D}}(k)$,

$$\tilde{W}_{3\mathrm{D}}(k) = \frac{\mathrm{d}}{\mathrm{d}\log k}W_{3\mathrm{D}}(k), \tag{75}$$

which for the 3D case can be written,

$$\tilde{W}_{3\mathrm{D}}(k) = \frac{1}{k^3}[(9k\cos(k) + 3(k^2-3)\sin(k)]. \tag{76}$$

Note that for a power law spectrum of index n_s this variance is only defined when $n_s < -1$. For practical application to cosmological models that resemble the concordant model, the effective index n_s decreases to -3 at small scales and the variance of $\hat{s}$ is always finite. This property however suggests that the amplitude of the slope fluctuations could be dominated by density fluctuations at scales significantly smaller than the smoothing radius if the latter is large enough. This is not expected to be the case however for the filtering scales we explore in this investigation. More precisely, provided the power spectrum index is in the range $[-3,-1]$, the amplitude of the variance of $\hat{s}$ can be expressed in terms of the variance of the density as

$$\langle\hat{s}^2\rangle = \sigma_R^2\frac{n_s(n_s+3)(n_s+5)}{4(n_s+1)}. \tag{77}$$

Let us now see how the whole statistical properties of the variable $\hat{s}$ can be derived from our formalism. Let us first explore the consequence of the change of variable, $(\hat{\rho}_1,\hat{\rho}_2) \to (\hat{\rho},\hat{s})$. Instead of describing the joint PDF as a function of the associated variables λ_1 and λ_2 we can build it with the variable associated to $\hat{\rho}$ and $\hat{s}$. Noting that $\lambda_1\hat{\rho}_1 + \lambda_2\hat{\rho}_2$ can be written as

$$\lambda_1\hat{\rho}_1 + \lambda_2\hat{\rho}_2 = (\lambda_1+\lambda_2)\hat{\rho}_1 + \frac{\Delta R}{R_1}\lambda_2\hat{s}, \tag{78}$$

as a consequence, the joint cumulant generating function of $\hat{\rho}_1$ and $\hat{s}$ is given by $\varphi(\lambda_1,\lambda_2)$ when written as a function of

$$\lambda = \lambda_1 + \lambda_2, \qquad \mu = \frac{\Delta R}{R_1}\lambda_2, \tag{79}$$

which are the variables associated with the Laplace and inverse Laplace transform of $\mathcal{P}(\hat{\rho}_1,\hat{s})$. One can also check that, following this definition, $\varphi(\lambda,\mu)$ is the Legendre transform of $\Psi(\rho_1, s = (\rho_2-\rho_1)R_1/\Delta R)$.

Let us then explore the whole statistical properties of $\hat{s}$ in the limit of a vanishing radius difference $(\Delta R)/R \to 0$. First note that the reduced skewness of $\hat{s}$ is still finite [62] and has a nontrivial value. It is given by

$$S_{03}^{\Delta R\to 0}=2+\left.\frac{\frac{\partial}{\partial R_1}\sigma^2_{R_1R_2}}{\frac{R_1\partial^2}{\partial R_1\partial R_2}\sigma^2_{R_1R_2}}\right|_{R=R_1=R_2}(6\nu_2-(\tilde{n}+3)), \quad (80)$$

where the effective index, $\tilde{n}$, is defined as

$$\frac{1}{\langle\hat{s}^2\rangle}\frac{\mathrm{d}}{\mathrm{d}\log R}\langle\hat{s}^2\rangle=-(\tilde{n}+3). \quad (81)$$

We will see in the following that this feature, the fact that reduced cumulants remain finite, extends to the whole generating function.

C. Analytic properties of $\varphi(\lambda,\mu)$

Let us now turn to the full analytical properties of $\varphi(\lambda,\mu)$, for a finite radius difference to start with, and then in the limit of vanishing radius difference. It is to be noted that, as for the one-cell case, not all values of λ and μ are accessible. This is due to the fact that the $\rho_i - \lambda_i$ relation cannot always be inverted via Eq. (24). The boundary of the region of interest is signaled by the fact that the determinant of the transformation vanishes, i.e., $\det[\partial^2\Psi(\{\rho_k\})/\partial\rho_i\partial\rho_j]=0$. This condition is met for finite values of both ρ_i and λ_i. The resulting critical lines are shown as thick solid lines in Fig. 5. Note that $\varphi(\lambda_1,\lambda_2)$ is also finite at this location. Within this line φ is defined; beyond this line it is not. Let us now explore the behavior of $\varphi(\lambda,\mu)$ when $\Delta R/R\to 0$. This is actually a cumbersome limit to take. One of the reasons is that the matrix Ξ_{ij} then becomes singular. More precisely the determinant of the cross-correlation function takes the form

$$\det[\Sigma_{ij}(R,R+\Delta R)]=R^{-2(3+n_s)}\left(\frac{\Delta R}{R}\right)^2\frac{-9+n_s^2}{4(1+n_s)}$$

at leading order in $\Delta R/R$ and when $n_s<-1$. For a power law spectrum, the actual coefficients read

$$\Xi_{11}(R,\Delta R)=\frac{2(n_s+1)R^{n_s+3}}{(\frac{\Delta R}{R})^2(n_s^2-9)}\left(\left(\frac{\Delta R}{R}\right)^2(n_s^2+7n_s+12)-2\frac{\Delta R}{R}(n_s+3)+2\right), \quad (82)$$

$$\Xi_{12}(R,\Delta R)=-\frac{R^{n_s+3}}{2(\frac{\Delta R}{R})^2(n_s^2-9)}\left(\left(\frac{\Delta R}{R}\right)^2(n_s^3+8n_s^2+23n_s+24)-4\frac{\Delta R}{R}(n_s^2+4n_s+3)+8(n_s+1)\right), \quad (83)$$

$$\Xi_{22}(R,\Delta R)=\frac{4(n_s+1)R^{n_s+3}}{(\frac{\Delta R}{R})^2(n_s^2-9)}. \quad (84)$$

All these coefficients are diverging like $(R/\Delta R)^2$. What we need to compute is however $\Psi(\rho,s)$ for finite values of ρ and s. In this case ρ_2 is also infinitely close to ρ_1 with $\rho_2-\rho_1=s\Delta R/R$ with a fixed value for s. Then the resulting value of $\Psi(\rho,s)$ is finite in the limit $\Delta R\to 0$. Assuming the form (13) for $\zeta(\tau)$ one gets

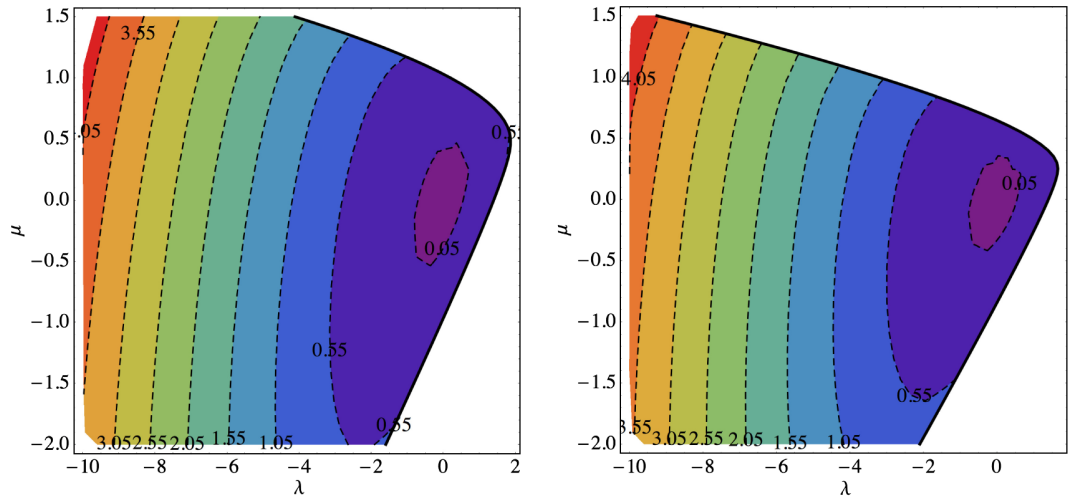


FIG. 5 (color online). Contour plot of $\varphi(\lambda,\mu)-\lambda$ (left panel) with a finite radius difference $\Delta R/R=1/10$ and (right panel) with $\Delta R/R\to 0$. We see that the structure of the critical region, although deformed, is preserved. In both cases, the restriction of $\varphi(\lambda,\mu)$ to $\mu=0$ is precisely the one-cell cumulant generating function considered in Sec. III.

$$\Psi(\rho,s)=\frac{R^{3+n_s}\rho^{n_s/3+(\nu-2)/\nu}}{2(n_s^2-9)(s+3\rho)^2}\{s^2[\nu^2n_s^3(\rho^{\frac{1}{\nu}}-1)^2+3n_s(5\nu^2(\rho^{\frac{1}{\nu}}-1)^2+16\nu(\rho^{\frac{1}{\nu}}-1)+12)$$
$$+4\nu n_s^2(\rho^{\frac{1}{\nu}}-1)(2\nu(\rho^{\frac{1}{\nu}}-1)+3)+36(\nu(\rho^{\frac{1}{\nu}}-1)+1)]+9\nu^2\rho^2n_s(n_s^2+8n_s+15)(\rho^{\frac{1}{\nu}}-1)^2$$
$$+6s\nu\rho(n_s+3)(\rho^{\frac{1}{\nu}}-1)(\nu n_s^2(\rho^{\frac{1}{\nu}}-1)+n_s(5\nu(\rho^{\frac{1}{\nu}}-1)+6)+6)\}. \quad (85)$$

The function $\varphi(\lambda,\mu)$ can then be obtained by Legendre transform. Like for the one-cell case, the transformation becomes critical when the inversion of the stationary condition is singular. For the new variables, it is also occurring when the determinant of the second derivatives of Ψ vanishes,

$$\det\left[\frac{\partial^2\Psi(\rho,s)}{\partial\rho\partial s}\right]=0, \quad (86)$$

which generalizes the condition (42). This condition defines the location of the critical line which can then be visualized in the $\lambda-\mu$ plane (thick lines in Fig. 5). Note that the no-shell crossing condition, which in this limit reads $s>-3\rho$, is located beyond this critical line and is therefore not relevant.

In the regular region, the contour lines of $\varphi(\lambda,\mu)$ are shown in Fig. 5 for both a finite ratio $\Delta R/R$ and when it is infinitely small. This figure explicitly shows in particular that the limit $\Delta R\to 0$ is nonpathological, in the sense that the location of the critical line and the actual value of the cumulant generating function converge to well-defined values in that limit. The convergence is however not very rapid and in practice we will use finite differences for comparisons with simulations.

Finally, to conclude this subsection we also compare these contour plots with those measured in simulations. There, one actually computes the explicit sum

$$\exp[\varphi(\lambda,\mu)]=\frac{1}{N_x}\sum_x\exp(\lambda\hat{\rho}_x+\mu\hat{s}_x), \quad (87)$$

where $\hat{\rho}_x$ and $\hat{s}_x$ are the measured values of $\hat{\rho}$ and $\hat{s}$ in a cell centered on x (in practice on grid points) and N_x is the number of points used (see Appendix D for details). Then $\varphi(\lambda,\mu)$ is always well defined, irrespective of the values of λ and μ. To detect the location of a critical line one should then rely on the properties it is associated with. From the analysis of the one-cell case it appears that for $\lambda>\lambda_c$, $\varphi(\lambda)$ is ill defined because $\int\mathcal{P}(\hat{\rho})\exp(\lambda\hat{\rho})d\hat{\rho}$ diverges. More precisely when $\lambda\to\lambda_c$ the value of $\varphi(\lambda)$ becomes dominated by the rare event tail. It makes such a quantity very sensitive to cosmic variance and in practice the critical line position is therefore associated with a diverging cosmic variance. In the two-cell case, we encounter the same effects. To locate we therefore simply cut out part of the $(\lambda-\mu)$ plane for which the measured variance of $\varphi(\lambda,\mu)$ exceeds a significant fraction of its measured value. We set this fraction to be 20% [63]. This criterium give rises to the solid line shown in Fig. 6. This figure is now to be compared to the left panel of Fig. 5. Although the figures are not identical they clearly exhibit the same patterns.

D. Slope cumulant generating function and PDF

When one wishes to build the PDF of $\hat{s}$, one needs to restrict $\varphi(\lambda,\mu)$ presented in the previous section to the $\lambda=0$ axis, i.e. focus on $\varphi_s(\mu)\equiv\varphi(\lambda=0,\mu)$. Figure 6 shows that $\varphi_s(\mu)$ has two extrema points, one corresponding to a positive value of μ, μ_c^+, and one to a negative value μ_c^-. The resulting global shape of $\varphi_s(\mu)$ is shown in Fig. 7, where it is also compared to the results where $\Delta R/R$ is kept finite. It actually shows that the limit $\Delta R/R$ is genuine at the level of the cumulant generating function but is reached for very small values of $\Delta R/R$. When predictions are compared with simulations for which the slope is measured with finite differences, it is necessary to use a finite difference ΔR.

We are now in position to build the one-point PDF of the density slope via the inverse Laplace transform of the cumulant generating function. It should be clear from the singular behavior of $\varphi_s(\mu)$ that it will exhibit exponential cutoffs on both sides, for positive and negative values of $\hat{s}$ although not *a priori* in a symmetric way. In practice, to do the complex plane integration, we build the function $\varphi_s(\mu)$ for the actual power spectrum of interest, and *then* build an

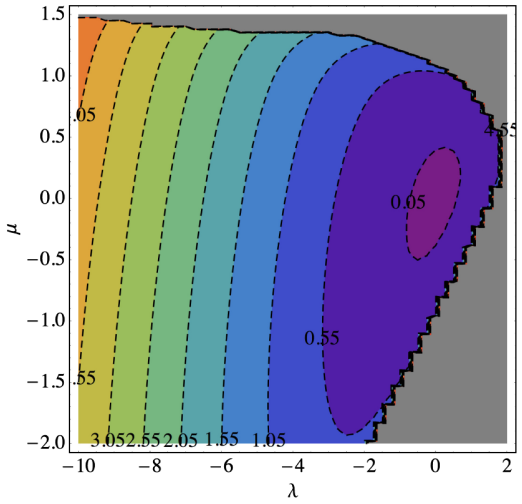


FIG. 6 (color online). Contour plot of $\varphi(\lambda,\mu)-\lambda$, from the simulation. It is to be compared with the left panel of Fig. 5.

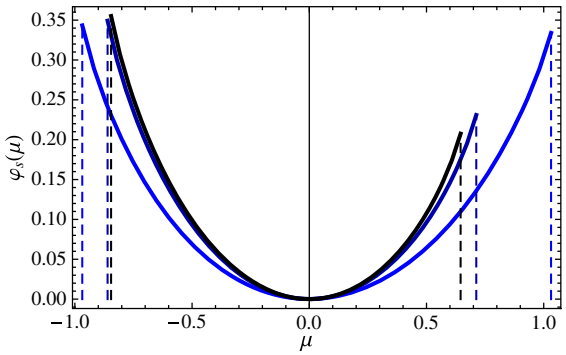


FIG. 7 (color online). The slope generating function $\varphi_s(\mu)$ for finite differences $\Delta R/R = 0.1$ and $\Delta R/R = 0.01$, and in the limit $\Delta R/R \to 0$. The corresponding curves are respectively in blue, darker blue and black. The vertical dashed lines show the locations of the critical points, μ_c^- and μ_c^+.

effective form $\zeta_{\rm eff}(\tau)$ that reproduces the numerical integration following Eqs. (38)–(39) as explained in [28]. In practice we use a seventh order polynomial to do the fit. We then proceed via integration in the complex plane using the usual approach (see Appendix B). The results for $R = 10h^{-1}$ Mpc and $z = 1.46$ and $z = 0.97$ is presented in the top panel of Fig. 8. The figure clearly exhibits the expected double cutoffs. Discrepancies between numerical results and theory that can be seen in the bottom panels for $\hat{s} \approx -0.5$ are not clearly understood (cosmic variance, numerical artifacts?).

E. The expected constrained slope and profile

Let us finally move to the key result of this paper. In the previous subsection we built the marginal PDF of $\hat{s}$; we now focus on the conditional properties of $\hat{s}$ given $\hat{\rho}_1 = \hat{\rho}(< R_1)$ at a given $R = R_1$, whether $\hat{s}$ is defined from a nearby radius of not. Mathematically it can be expressed in terms of the joint PDF, $\mathcal{P}(\hat{\rho}_1, \hat{\rho}_2)$, as

$$\langle \hat{s} \rangle_{\hat{\rho}_1} = -\frac{R}{\Delta R}\hat{\rho}_1 + \frac{R}{\Delta R \mathcal{P}(\hat{\rho}_1)} \int \mathrm{d}\hat{\rho}_2 \hat{\rho}_2 \mathcal{P}(\hat{\rho}_1, \hat{\rho}_2), \quad (88)$$

given that

$$\int \mathrm{d}\hat{\rho}_2 \hat{\rho}_2 \mathcal{P}(\hat{\rho}_1, \hat{\rho}_2) = \int_{-i\infty}^{+i\infty} \frac{\mathrm{d}\lambda_1}{2\pi i} \frac{\partial \varphi(\lambda_1, \lambda_2)}{\partial \lambda_2}\bigg|_{\lambda_2=0} \exp(-\lambda_1 \hat{\rho}_1 + \varphi(\lambda_1)), \quad (89)$$

which can be obtained by explicit integration in the complex plane [64]. Note that the solution of the stationary equations, Eq. (28), yields the identity

$$\frac{\partial \varphi(\lambda_1, \lambda_2)}{\partial \lambda_2}\Big|_{\lambda_2=0} = \rho_2(\lambda_1, \lambda_2 = 0). \quad (90)$$

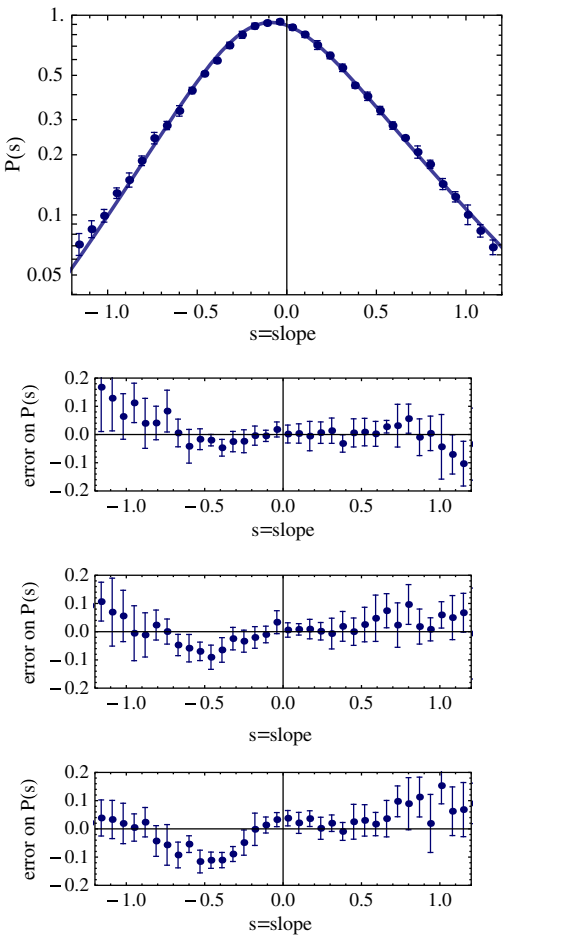


FIG. 8 (color online). The PDF of the slope for $z = 1.46$. The bottom panels show the residuals for $z = 1.46$, $z = 0.97$ and $z = 0.65$ from top to bottom.

For the saddle point solution corresponding to the low ρ regime, λ_1 and $\hat{\rho}_1$ in Eq. (89) are related through the stationary condition. In this limit we therefore have

$$\langle \hat{\rho}_2 \rangle_{\hat{\rho}_1} = \bar{\rho}_2(\hat{\rho}_1), \quad (91)$$

where $\bar{\rho}_2(\hat{\rho}_1)$ is the solution of the system

$$\lambda_1 = \frac{\partial \Psi(\rho_1, \bar{\rho}_2)}{\partial \rho_1}, \qquad 0 = \frac{\partial \Psi(\rho_1, \bar{\rho}_2)}{\partial \rho_2}. \quad (92)$$

These calculations can be extended to the constrained variance of the slope. The computation follows the same line of derivation but is slightly more involved. It is presented in Appendix C.

Let us now present the expected slope from exact complex plane numerical integration, using the analytical saddle point approximations and as measured in numerical simulations. For instance, Fig. 9 shows the expected slope given by $10. \times [\hat{\rho}(1.1R) - \hat{\rho}(R)]$ as a function of $\hat{\rho}(R)$

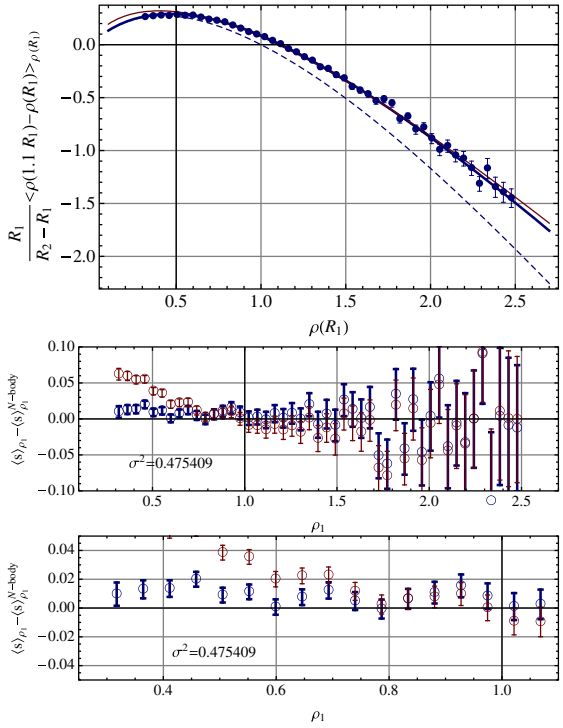


FIG. 9 (color online). (Top panel) The conditional profile, $\langle\hat{s}\rangle_{\hat{\rho}(<R_1)}$ as a function of $\hat{\rho}(<R_1)$. The thick blue solid line is the result of the numerical integration, the thin dashed line the saddle point approximation Eq. (91). We also present the power law approximation case as a thin (red) solid line. It is shown to depart from the exact prediction in the low density region. The agreement between the theory and the measurements near the origin is quite remarkable. The bottom panels show the residuals computed in bins as a function of the density (with a zoomed plot below). Again the thick symbols correspond to the exact calculation, the thin symbols correspond to the power law approximation.

using the same cosmological parameters as for Fig. 2. The solid lines are the results of complex plane integrations and the dashed line is the saddle point approximation. The latter is found to perform very poorly when compared to simulation. We note also that the low density part of the prediction can only be accounted for when the running parameter is taken into account. This is clearly visible in the middle and bottom panels when one compares the (thick) blue and the (thin) red marks. These comparisons show that the analytical predictions are accurate at percent level in a large range of parameters.

Let us finally turn to the more global properties of the density in cells and consider the density *profile* defined as the constrained density $\hat{\rho}(R_2)$ given $\hat{\rho}(R_1)$ as a function of R_2. Technically computing expected profiles or slopes is equivalent. The second point of view allows however to visualize what should be the radial variation of the density profile, and its fluctuations, of an underdense or an overdense region. The result of such a calculation is presented

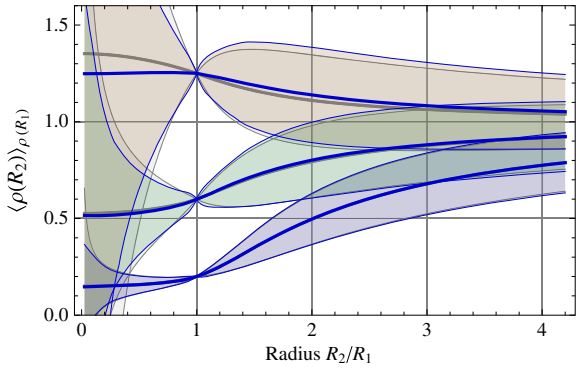


FIG. 10 (color online). The conditional profile as a function of R_2 and for different choices of $\hat{\rho}(R_1)$ [to which $\hat{\rho}(R_2)$ is equal to at $R_2 = R_1$]. The blue thick solid lines are the results of numerical integrations; the colored thin lines show the 1-σ variance about the expectation. The close-by gray lines are the same calculations but using the saddle point approximations of Eqs. (91) and (C5) respectively. Note in particular the smaller variance of the underdense profile near $R_2 \sim 0$.

in Fig. 10 which shows the expected density as a function of the radius R_2 and for various values of $\hat{\rho}(R_1)$. In the same plot we also show the expected 1-σ variance about the expectation values. Both quantities are computed using the exact complex plane integration and compared to their saddle point approximation counterparts. The difference is only significant for $\hat{\rho}(R_1) = 1.25$. Interestingly for low density prior, e.g. $\hat{\rho}(R_1) = 0.2$ in the figure, the variance is small (and significantly smaller than the variance of $\hat{s}$ in the absence of prior on the density). That implies that all voids should look similar, probably a good starting point for exploring the statistical properties of the field while focussing on these regions.

Comparisons of the latter prediction with numerical simulations is made in Fig. 11 where we give both the measurements of the expected profile and their variance for a given constraint. The only difference with the theoretical predictions is that the constraints are binned, i.e. the prior is that the density ρ_1 is assigned in a given bin of width 0.2 centered on the values 0.35,0.74,1.13,1.52,1.92. As the theoretical predictions do not take into account the binning, there is a noticeable departure between the predicted variance and its measured value near $R_2/R_1 \approx 1$ due to the width of the bin. But, this departure notwithstanding, the agreement between the theoretical predictions and the measured quantities, for both the expected profile and its variance, is just striking. Only when the constraint density is large (top two panels) can we see some slight departure of the variances with the theory for small radii, which is due to the fact that they correspond to regions entering the nonlinear regime.

F. Joint *n*-cells PDF

The results presented in the previous section give us confidence in the general framework we have adopted here.

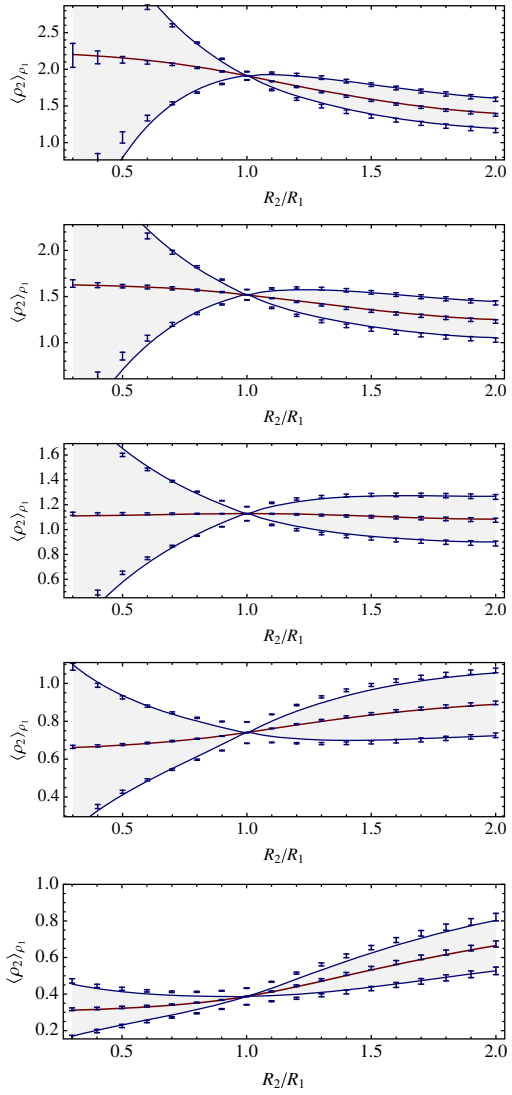


FIG. 11 (color online). Same quantities as in Fig. 10 measured here in simulations. The solid lines are the theoretical predictions and the points with error bars are the measurements for both the expected value and its variance. The agreement is spectacular in particular for low density constraints.

It has to be stressed that all of the properties we have described are simultaneously captured with the shape of the multiple cell cumulant generating function, $\varphi(\{\lambda_k\})$, or its counterpart, the n-cell PDF, $\mathcal{P}(\{\hat{\rho}_k\})$. We should keep in mind that we could have considered the former only to compare with simulations but we dramatically lack intuition for such a representation. By contrast we have much better intuition of what n-cell PDFs are. So far we have considered only the one-cell PDF. In the following we succinctly consider the derivation of the multicell PDF in our framework.

Hence let us consider a set of n concentric cells and its cumulant generating function $\varphi(\{\lambda_k\})$. In principle the corresponding PDF, $\mathcal{P}(\{\hat{\rho}_k\})$, is to be obtained from inverse Laplace transform. Such a computation appears extremely challenging to implement and we have not succeeded yet in producing a full two-cell PDF. We can however present its low density approximation, the counterpart of Eq. (46), for a multidimensional case. It is based on the use of the saddle point approximation of Eq. (40) assuming the overall variance is small. It leads to a similar condition that should be met at the saddle point $\{\lambda_s\}_i$,

$$\frac{\partial}{\partial\lambda_k}\Big[\sum_i \lambda_i\hat{\rho}_i - \varphi(\{\lambda_i\})\Big] = 0, \tag{93}$$

which leads to

$$\hat{\rho}_i = \rho_i(\{\lambda_k\}), \tag{94}$$

and with the constraint that

$$\det\left[\frac{\partial^2\Psi}{\partial\rho_k\partial\rho_l}\right] > 0 \tag{95}$$

at the saddle point position. The resulting expression for the density PDF generalizes Eq. (46) to

$$\mathcal{P}(\{\hat{\rho}_k\}) = \frac{1}{(2\pi)^{n/2}}\sqrt{\det\left[\frac{\partial^2\Psi}{\partial\hat{\rho}_k\partial\hat{\rho}_l}\right]}\exp\left[-\Psi(\{\hat{\rho}_k\})\right]. \tag{96}$$

This analytic expression is expected to be an approximate form for the exact PDF in underdense regions.

We suggest that in the absence of computable multiple cell PDFs, this form could be used, provided one makes sure to restrict its application to its proper region of validity. It is interesting to note that, in this framework, the *parameter dependence* of the mode of that PDF and its local curvature tensor can be straightforwardly computed from it analytically. In the concluding section we will simply sketch a way to constrain key cosmological parameters using this form.

V. CONCLUSIONS AND PROSPECTS

A. Summary

In the context of upcoming large wide field surveys we revisited the derivation of the cumulant generating functions of densities in spherical concentric cells in the limit of a vanishing variance and we conjectured that it correctly represents the generating function for finite values of the variance. We noted that such a quantity is an observable in itself and could probably be used as a cosmological indicator. In this study we however focused our efforts on its counterpart, the multicell density PDF.

We first computed the resulting one-cell density PDF. These results were tested with unprecedented accuracy, in

particular taking into account the scale variation of the power spectrum index. Comparisons to modern N-body simulations showed that predictions reach percent order accuracy (when the density variance is measured from simulations) for a large range of density values, as long as the variance is small enough. It confirmed in particular that this formalism gives a good account of the rare event tails: predictions are in agreement with the numerical measurements down to numerical precision.

We took advantage of the finite variance generating function formalism to explore its implications to the two-cell case in a novel regime. In particular we derived the statistical properties of the local density slope, defined as the infinitesimal difference of the density in two concentric cells of (possibly infinitesimally) close radii. We gave its mean expectation, and its expectation constrained to a given density. From the properties of the local slope, one can also construct the overall expected profile, i.e. the density as a function of the radius, and its fluctuations. We found the latter to be of particular interest when focusing on voids, as in these regions, the variances around the mean profile are significantly reduced (though the relative fluctuations less so). In particular we suggest below a possible method to constrain cosmological and gravity models from these low density regions. All these predictions were successfully compared to simulations.

B. Prospects

The full statistical power of the approach presented in this paper would ultimately be encoded in the shape of the two-cell density PDF but we do not know at this stage how to properly invert the exact expression given by Eq. (40) in this two-cell regime. Despite this limitation, as we do not have simulations that span different gravity models, let us use the saddle point form of Eq. (96) assuming it is exact (hence avoiding the issue of the domain of validity of that analytic fit to the exact PDF), and use its dependence on key (cosmic) parameters to infer the precision with which cosmological parameters could be constrained.

Focusing the analysis on two quantities, the parameter ν that encodes the spherical collapse dynamics [see Eq. (13)] and the power law index n_s, let us simply consider sets of about 2 000 and 11 000 independent measurements drawn in concentric spheres of radii 10 and 11 Mpc$/h$ and such that $0.05 < \rho_1 < 0.5$ and $-0.02 < \Delta\rho/\rho_1 < 0.06$ (i.e. near the peak of the PDF). The samples are drawn directly from the two-cell PDF for chosen values of the power law power spectrum and ν parameter starting respectively with $n_s = -2.5$ and $\nu = 3/2$. The likelihood of the models where n_s and ν vary in the range $[-0.12, 0.12]$ around the reference value is computed.

The resulting mean (over 25 independent samples) log likelihood of the data set as a function of $\Delta\nu/\nu$ and $\Delta n_s/n_s$ is displayed in Fig. 12 (at one, three and five sigma respectively) [65]. As expected, the likelihood contours

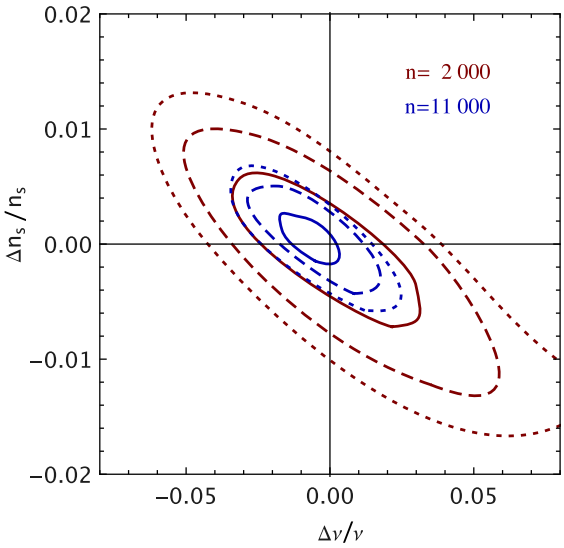


FIG. 12 (color online). The likelihood contours at one, three and five sigmas around the reference model $(n_s = -2.5, \nu = 3/2)$, drawn from ∼11 000 measurements (inner blue contours) and ∼2 000 measurements (outer red contours) of the densities in concentric shells of radii 10 and 11 Mpc$/h$.

are centered on the zero offset values; they yield the precision that could be reached in a survey of a useful volume of about $(200h^{-1}\ \mathrm{Mpc})^3$ (red contours) and $(360h^{-1}\ \mathrm{Mpc})^3$ (blue contours). These sample sizes are not unreasonable. Indeed, the volume span with about 30 000 spheres corresponds to the volume covered by the simulation and we found that, in doing so, the error bars on relevant quantities such as profiles (as shown in Fig. 9) were of the same order as the one measured in the simulation we used throughout the paper. At face value, relative accuracies below the percent on (n_s, ν) could be reached with such surveys. Yet, this numerical experiment is, at this stage, at best illustrative. We are indeed aware that in more realistic situations, one would have to properly account for the domain of validity of the above functional form, which would take us beyond the scope of this paper. Another open question would be to estimate how many concentric cells should be used to get an optimal constraint for a given set of cosmic parameters, but the answer to this question will probably depend on the geometry of the available survey.

Should these problems be alleviated, effective implementation of such cosmological tests would still be far fetched. In particular galaxy catalogs in z space break the local spherical symmetry in a complex way making the application of such a method impractical. One way to avoid this problem is to stick to observations for which this method is applicable, such as *projected* densities along the line of sight. It can be done either in the context of cosmic shear observations or for photometriclike redshift surveys. In both cases the point is not to reconstruct the spherical 3D

statistics but the circular two-dimensional statistics for which the whole method should be applicable following early investigations in [27,28]. The accuracy of the predictions still has to be assessed in this context. Another missing piece that can be incorporated is the large distance correlation of statistical indicators such as profiles and constrained profiles. Following [27] it is indeed within reach of this formalism to compute such quantities. We would then have a fully working theory that could be exploited in real data sets.

ACKNOWLEDGMENTS

We warmly thank D. Pogosyan for triggering our interest in studying the statistics of void regions. We also thank him for his many comments. This work is partially supported by Grants No. ANR-12-BS05-0002 and No. ANR-13-BS05-0005 of the French Agence Nationale de la Recherche and by the National Science Foundation under Grant No. NSF PHY11-25915. The simulations were run on the HORIZON cluster. C. P. thanks KITP and the University of Cambridge for their hospitality while this work was completed. We acknowledge support from S. Rouberol for running the cluster for us.

APPENDIX A: RADII DECIMATIONS

The purpose of this appendix is to make sure that the expression of $\varphi(\{\lambda\})$ is consistent with variable decimation, i.e. we want to make sure that

$$\varphi(\{\lambda_1,\ldots,\lambda_n\}) = \varphi(\{\lambda_1,\ldots,\lambda_n,\lambda_{n+1}=0,\ldots,\lambda_{n+m}=0\}), \quad \text{(A1)}$$

where the left-hand side is computed from n cells whereas the right-hand side is computed with $n+m$ cells.

In order to prove this property, let us define a set $\mathcal{A}$ of n cells and a set $\mathcal{B}$ of m cells. One can then define the covariance matrix $\sigma_{ij}(\rho_i,\rho_j)$ as in (17) between two any cells of the union of $\mathcal{A}$ and $\mathcal{B}$.

We first need to establish a preliminary relation between the element of the inverse matrix $\Xi_{ij}(\{\rho_k\})$ and the covariance matrix. From

$$\sum_{l=1}^{n}\sigma_{il}(\rho_i,\rho_l)\Xi_{lj}(\{\rho_k\}) = \delta_{ij}, \quad \text{(A2)}$$

we indeed can derive the following relation,

$$\sigma_{il}(\rho_i,\rho_l)\frac{\partial}{\partial\rho_k}[\Xi_{lj}(\{\rho_k\})]\sigma_{jm}(\rho_j,\rho_m) + \frac{\partial}{\partial\rho_k}[\sigma_{im}(\rho_i,\rho_m)] = 0, \quad \text{(A3)}$$

where all the repeated indices run from 1 to $n+m$. One can also write this relation when the inverse matrix is defined from the covariance matrix of the cells restricted in $\mathcal{A}$ only. Let us define by $\hat{\Xi}_{\mu\nu}(\{\rho_{\rho\mu}\})$ this matrix and in the following restrict the Greek indices from 1 to n. The previous relation is then transformed into

$$\sigma_{\mu\lambda}(\rho_\mu,\rho_\lambda)\frac{\partial}{\partial\rho_\kappa}[\hat{\Xi}_{\lambda\nu}(\{\rho_\mu\})]\sigma_{\nu\sigma}(\rho_\nu,\rho_\sigma) + \frac{\partial}{\partial\rho_\kappa}[\sigma_{\mu\sigma}(\rho_\mu,\rho_\sigma)] = 0. \quad \text{(A4)}$$

The cumulant generating functions for the n cells in $\mathcal{A}$ is given by

$$\hat{\varphi}(\{\lambda_\mu\}) = \lambda_\mu\tau_\mu - \frac{1}{2}\hat{\Xi}_{\mu\nu}\tau_\mu\tau_\nu, \quad \text{(A5)}$$

with the stationary conditions

$$\lambda_\kappa = \hat{\Xi}_{\mu\kappa}\tau_\mu\frac{\mathrm{d}\tau_\kappa}{\mathrm{d}\rho_\kappa} + \frac{1}{2}\frac{\partial\hat{\Xi}_{\mu\nu}}{\partial\rho_\kappa}\tau_\mu\tau_\nu. \quad \text{(A6)}$$

The purpose of the following calculation is to show that it is identical to the expression of $\varphi(\lambda_i)$ describing the cumulant generating function of the $n+m$ cells when the last $m-n$ values of λ_i are set to zero. In this case we have

$$\varphi(\{\lambda_\mu,0\}) = \lambda_\mu\tau_\mu - \frac{1}{2}\Xi_{ij}\tau_i\tau_j, \quad \text{(A7)}$$

with the stationary conditions

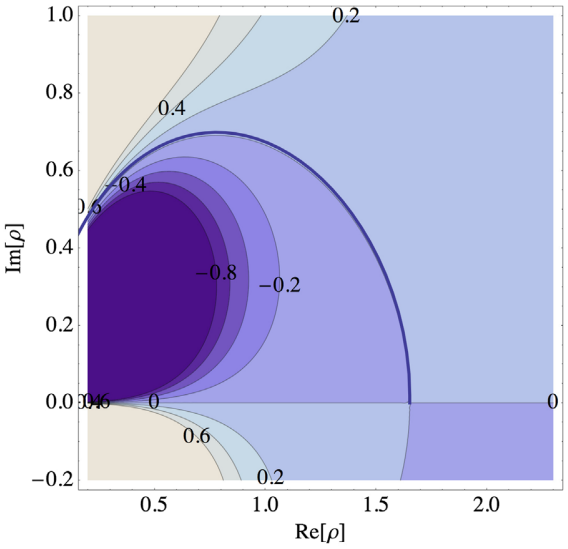


FIG. 13 (color online). The path line in the ρ complex plane. We superimposed the contour plot of the imaginary part of $\varphi(\lambda) - \lambda\hat{\rho}$ to check that it follows a $\Im[\varphi(\lambda) - \lambda\hat{\rho}] = 0$ line. The starting point on the real axis correspond to the saddle point value.

$$\lambda_\kappa = \Xi_{i\kappa}\tau_i \frac{\mathrm{d}\tau_\kappa}{\mathrm{d}\rho_\kappa} + \frac{1}{2}\frac{\partial \Xi_{ij}}{\partial \rho_\kappa}\tau_i\tau_j, \tag{A8}$$

$$0 = \Xi_{ki}\tau_i \frac{\mathrm{d}\tau_k}{\mathrm{d}\rho_k} + \frac{1}{2}\frac{\partial \Xi_{ij}}{\partial \rho_k}\tau_i\tau_j, \tag{A9}$$

for k running from $n+1$ to $n+m$. The second set of constraints allows to determine the values of τ_i for $i \in [n+1, n+m]$ in terms of τ_ν. It is given by

$$\hat{\tau}_i = \sigma_{i\mu}\hat{\Xi}_{\mu\nu}\tau_\nu, \tag{A10}$$

where once again repeated Greek indices are summed over from 1 to n. This expression is actually valid for any values of i as when i is in the 1 to n range we identically have $\hat{\tau}_i = \tau_i$. One can indeed check that for this expression the two terms in Eq. (A9) are identically 0: indeed $\Xi_{ki}\hat{\tau}_i = \delta_{k\mu} = 0$ for $k \in [n+1, n+m]$ and $\partial\Xi_{ij}/\partial\rho_k\hat{\tau}_i\hat{\tau}_j = \partial\Xi_{ij}/\partial\rho_k\sigma_{i\mu}\hat{\Xi}_{\mu\nu}\tau_\nu\sigma_{j\mu'}\hat{\Xi}_{\mu'\nu'}\tau_{\nu'} = -\partial\sigma_{\mu\mu'}/\partial\rho_k\hat{\Xi}_{\mu\nu}\tau_\nu\hat{\Xi}_{\mu'\nu'}\tau_{\nu'} = 0$ for $k \in [n+1, n+m]$. Then replacing using this expression for the τ_i in Eq. (A8), one gets

$$\lambda_\kappa = \Xi_{i\kappa}\sigma_{i\mu}\hat{\Xi}_{\mu\nu}\tau_\nu \frac{\mathrm{d}\tau_\kappa}{\mathrm{d}\rho_\kappa} + \frac{1}{2}\frac{\partial \Xi_{ij}}{\partial \rho_\kappa}\sigma_{i\mu}\hat{\Xi}_{\mu\nu}\tau_\nu\sigma_{j\mu'}\hat{\Xi}_{\mu'\nu'}\tau_{\nu'}.$$

Its first term can be simplified using the definition of $\hat{\Xi}$ and the second by the subsequent use of Eqs. (A3) and (A4),

$$\begin{aligned}\frac{\partial \Xi_{ij}}{\partial \rho_\kappa}\sigma_{i\mu}\hat{\Xi}_{\mu\nu}\sigma_{j\mu'}\hat{\Xi}_{\mu'\nu'} &= -\frac{\partial}{\partial \rho_\kappa}\sigma_{\mu\mu'}\hat{\Xi}_{\mu\nu}\hat{\Xi}_{\mu'\nu'}, \\ &= \frac{\partial \hat{\Xi}_{\kappa\sigma}}{\partial \rho_\kappa}\sigma_{\kappa\mu}\hat{\Xi}_{\mu\nu}\sigma_{\sigma\mu'}\hat{\Xi}_{\mu'\nu'}, \\ &= \frac{\partial \hat{\Xi}_{\nu\nu'}}{\partial \rho_\kappa}, \end{aligned} \tag{A11}$$

so that the expression of λ_κ coincides with the expression (A6). Finally $\hat{\tau}_\mu = \tau_\mu$ ensures that the property (A1) is valid.

APPENDIX B: INTEGRATION IN THE COMPLEX PLANE

1. Numerical algorithm

The computation of the one-point PDF relies on the following expression:

$$P(\hat{\rho}) = \int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} \frac{\mathrm{d}\lambda}{2\pi\mathrm{i}} \exp(-\lambda\hat{\rho} + \varphi(\lambda)), \tag{B1}$$

where we explicitly denote $\hat{\rho}$ as the value of the density for which we want to compute the PDF. This is to distinguish it from the variable ρ that enters in the calculation of $\varphi(\lambda)$ out of the Legendre transform of $\Psi(\rho)$. The idea to achieve fast convergence of the integral is to follow a path in the complex plane where the argument of the exponential in Eq. (B1) is real. The starting point of the calculation is $\rho = \rho_s$. When $\hat{\rho}$ is small enough (in the regular region) then we simply have $\rho_s = \hat{\rho}$ otherwise one should take $\rho_s = \rho_c$. At this very location, two lines of vanishing imaginary parts of $-\lambda\hat{\rho} + \varphi(\lambda)$ cross, one along the real axis (obviously) and one parallel to the imaginary axis (precisely because we are at a saddle point position). The idea is then to build, step by step, a path by imposing

$$\delta[\varphi(\lambda) - \lambda\hat{\rho}] \in \mathbb{R}. \tag{B2}$$

This condition can be written as an infinitesimal variation of λ. Recalling that $\mathrm{d}\varphi(\lambda)/\mathrm{d}\lambda = \rho(\lambda)$, for each step we have to impose

$$(\rho - \hat{\rho})\delta\lambda \in \mathbb{R}, \tag{B3}$$

which in turns can be obtained by imposing that the complex argument of $(\delta\rho)$ is that of $[(\rho - \hat{\rho})\mathrm{d}^2\Psi/\mathrm{d}\rho^2]^*$. This is what we implement in practice. Accurate prediction for the PDFs are obtained with about 50 points along the path line that is illustrated on Fig. 13.

2. The large density tails

The derivation of the rare event tail of the density PDF for large positive densities is based on the inverse Laplace transform of the generating function $\varphi(\lambda)$ when it is dominated by its singular part, i.e. for $\lambda \approx \lambda_c$. In this case the complex plane contour is pushed along the real axis wrapping around the singular value λ_c as depicted on Fig. 14.

The general form for the density PDF given by Eq. (B1) is expressed using the form (43) following the path shown on Fig. 14. As the contributions from the two branches of the path lines are complex conjugate, it eventually leads to the form

$$\begin{aligned}P(\hat{\rho}) \approx \Im\Big\{ &\int_{\mathrm{i}\epsilon+\lambda_c}^{\mathrm{i}\epsilon+\infty} \frac{\mathrm{d}\lambda}{\pi} \exp[\varphi_c - \lambda_c\hat{\rho} - (\lambda - \lambda_c)(\hat{\rho} - \rho_c)] \\ &\times [1 + a_{3/2}(\lambda - \lambda_c)^{3/2} + \cdots]\Big\}, \end{aligned} \tag{B4}$$

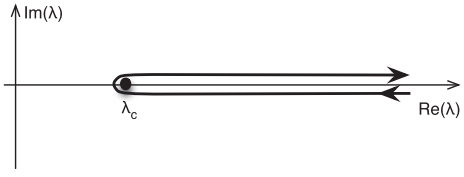


FIG. 14. The path line in the λ complex plane for the computation of the large density asymptotic forms.

where we keep only the dominant singular part in $\varphi(\lambda)$ and where $\Im$ denotes the imaginary part. This integral can easily be computed and it leads to

$$P(\hat{\rho}) \approx \exp(\varphi_c - \lambda_c\hat{\rho})\left(\frac{3\Im(a_{\frac{3}{2}})}{4\sqrt{\pi}(\hat{\rho}-\rho_c)^{5/2}} + \cdots\right). \quad \text{(B5)}$$

Subleading contributions can be computed in a similar way when $\exp(\varphi(\lambda))$ is expanded to higher order. Note that by symmetry, only half integer terms that appear in this expansion will actually contribute.

APPENDIX C: THE CONSTRAINED VARIANCE

In this appendix we complement the calculations started in Sec. IV E where we computed the expected slope under a local density constraint. Pursuing along the same line of calculations, the variance of $\hat{\rho}_2$ given $\hat{\rho}_1$ can be computed from the conditional value of $\hat{\rho}_2^2$. It is given by the second derivative of the moment generating function, and is therefore given by

$$\int \mathrm{d}\hat{\rho}_2\hat{\rho}_2^2 P(\hat{\rho}_1,\hat{\rho}_2) = \int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\frac{\mathrm{d}\lambda_1}{2\pi\mathrm{i}}\left[\frac{\partial^2\varphi(\lambda_1,\lambda_2)}{\partial\lambda_2^2}\bigg|_{\lambda_2=0} + \left(\frac{\partial\varphi(\lambda_1,\lambda_2)}{\partial\lambda_2}\bigg|_{\lambda_2=0}\right)^2\right]\exp(-\lambda_1\hat{\rho}_1+\varphi(\lambda_1)).$$

The calculation of its approximate form in the low-ρ saddle point limit is a bit more cumbersome. Indeed, in the low variance limit in which this approximation is derived the two terms in the square brackets are not of the same order, the first being subdominant with respect the second. It is nonetheless possible to compute the resulting cumulant in the low density limit. Formally, differentiating Eq. (90) with respect to λ_2 we have

$$\frac{\partial^2\varphi(\lambda_1,\lambda_2)}{\partial\lambda_2^2} = \frac{\partial\rho_2(\lambda_1,\lambda_2)}{\partial\lambda_2}, \quad \text{(C1)}$$

from the Legendre stationary condition, which, after inversion of the partial derivatives, is formally given by

$$\frac{\partial^2\varphi(\lambda_1,\lambda_2)}{\partial\lambda_2^2} = \frac{\Psi_{,\rho_1\rho_1}}{\Psi_{,\rho_1\rho_1}\Psi_{,\rho_2\rho_2}-\Psi^2_{,\rho_1\rho_2}}, \quad \text{(C2)}$$

where $\Psi_{,\rho_i\rho_j} \equiv \partial^2\Psi/\partial\rho_i\partial\rho_j$ are calculated at the stationary point. On the other hand $\partial\varphi(\lambda_1,\lambda_2)/\partial\lambda_2$ can be expanded as

$$\frac{\partial\varphi(\lambda_1,\lambda_2)}{\partial\lambda_2} = \varphi(\lambda_s,0) + (\lambda_1-\lambda_s)\times\frac{\partial^2\varphi(\lambda_1,\lambda_2)}{\partial\lambda_2\partial\lambda_1} + \frac{1}{2}(\lambda_1-\lambda_s)^2\frac{\partial^3\varphi(\lambda_1,\lambda_2)}{\partial\lambda_2\partial\lambda_1^2}+\cdots \quad \text{(C3)}$$

near the saddle point value λ_s. The integration of $\partial\varphi(\lambda_1,\lambda_2)/\partial\lambda_2$ in the complex plane therefore leads to a correction from the $(\lambda_1-\lambda_s)^2$ term. It can be verified though that this contribution vanishes when one takes the cumulant. The integration of $(\partial\varphi(\lambda_1,\lambda_2)/\partial\lambda_2)$ however leads to an extra term due to the second term in the previous expansion. The resulting term reads $[\partial^2\varphi/\partial\lambda_2\partial\lambda_1]^2/\partial^2\varphi/\partial\lambda_1^2$, so that

$$\langle\rho_2^2\rangle_{\rho_1} - \langle\rho_2\rangle^2_{\rho_1} = \frac{\partial^2\varphi}{\partial\lambda_1^2} - \left[\frac{\partial^2\varphi}{\partial\lambda_2\partial\lambda_1}\right]^2\Big/\frac{\partial^2\varphi}{\partial\lambda_1^2} = \left[\frac{\partial^2\varphi}{\partial\lambda_1^2}\frac{\partial^2\varphi}{\partial\lambda_2^2} - \left(\frac{\partial^2\varphi}{\partial\lambda_1\partial\lambda_2}\right)^2\right]\Big/\frac{\partial^2\varphi}{\partial\lambda_1^2}, \quad \text{(C4)}$$

which can be rewritten more compactly as

$$\langle\hat{\rho}_2^2\rangle_{\hat{\rho}_1} - \langle\hat{\rho}_2\rangle^2_{\hat{\rho}_1} = 1/\Psi_{,\rho_2\rho_2}|_{\hat{\rho}_1,\bar{\rho}_2(\hat{\rho}_1)}, \quad \text{(C5)}$$

when expressed in terms of Ψ.

APPENDIX D: SIMULATIONS

For the purpose of this paper, we have carried out a dark matter simulation with GADGET2 [66]. This simulation is characterized by the following ΛCDM cosmology: $\Omega_\mathrm{m} = 0.265$, $\Omega_\Lambda = 0.735$, $n = 0.958$, $H_0 = 70$ km × s^{-1} × Mpc^{-1} and $\sigma_8 = 0.8$, $\Omega_b = 0.045$ within one standard deviation of Wilkinson Microwave Anisotropy Probe 7 results [67]. The box size is 500 Mpc/h sampled with 1024^3 particles, the softening length 24 kpc/h. Initial conditions are generated using MPGRAFIC [68]. The variances and running indexes are measured from the theoretical power spectra produced by MPGRAFIC. Snapshots are saved for $z = 0, 0.65, 0.97, 1.46, 2.33$ and 3.9. An Octree is built for each snapshot, which allows us to count very efficiently all particles within a given sequence of concentric spheres of radii between $R = 4, 5\ldots$ up to 18 Mpc/h. The center of these spheres is sampled regularly on a grid of 10 Mpc/h aside, leading to 117 649 estimates of the density per snapshot. All histograms drawn in this paper are derived from these samples. Note that the cells overlap for radii larger than 10 Mpc/h.

APPENDIX E: SYSTEMATIC COMPARISONS WITH SIMULATION

We collect here figures that are too large to be put in the main text.

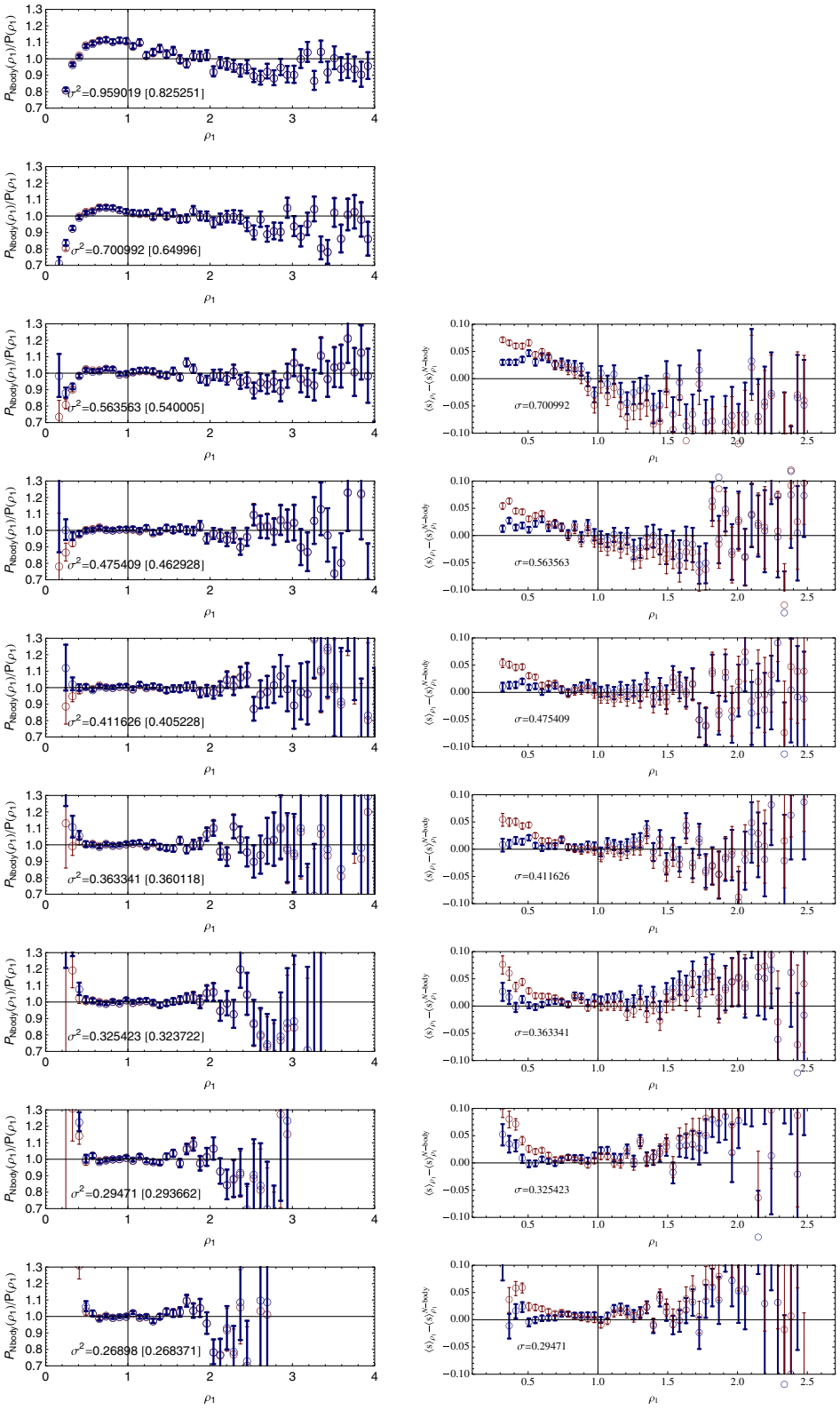


FIG. 15 (color online). (Left panels) The residuals of the expected density PDF from smoothing scale of $R_1 = 4$ (top panel) to $R_1 = 20h^{-1}$ Mpc. This is for $z = 0.97$ (same convention as in Fig. 3). The values of σ^2 at the smoothing scale is given in the inset. The value in square brackets is the linear value. (Right panels) The residuals for the expected average profile between scale R_1 and $R_2 = R_1 + 1h^{-1}$ Mpc. From top to bottom we have $R_1 = 6$ to $18h^{-1}$ Mpc. Same convention as in Fig. 8.

[1] See http://bigboss.lbl.gov, https://www.darkenergysurvey.org, http://pan-starrs.ifa.hawaii.edu, and http://www.lsst.org.
[2] R. Laureijs, J. Amiaux, S. Arduini, J.-L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, arXiv:1110.3193.
[3] See http://sci.esa.int/euclid, http://snap.lbl.gov, and http://jdem.lbl.gov.
[4] S. F. Daniel, E. V. Linder, T. L. Smith, R. R. Caldwell, A. Cooray, A. Leauthaud, and L. Lombriser, Phys. Rev. D **81**, 123508 (2010).
[5] A. F. Heavens, T. D. Kitching, and L. Verde, Mon. Not. R. Astron. Soc. **380**, 1029 (2007).
[6] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002).
[7] M. Crocce and R. Scoccimarro, Phys. Rev. D **73**, 063519 (2006).
[8] A. Taruya and T. Hiramatsu, Astrophys. J. **674**, 617 (2008).
[9] M. Pietroni, J. Cosmol. Astropart. Phys. 10 (2008) 036.
[10] M. Crocce, R. Scoccimarro, and F. Bernardeau, Mon. Not. R. Astron. Soc. **427**, 2537 (2012).
[11] A. Taruya, F. Bernardeau, T. Nishimichi, and S. Codis, Phys. Rev. D **86**, 103528 (2012).
[12] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, J. High Energy Phys. 09 (2012) 082.
[13] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[14] J. R. Gott III, Publ. Astron. Soc. Pac. **100**, 1307 (1988).
[15] K. R. Mecke, T. Buchert, and H. Wagner, Astron. Astrophys. **288**, 697 (1994).
[16] S. Colombi, D. Pogosyan, and T. Souradeep, Phys. Rev. Lett. **85**, 5515 (2000).
[17] C. Gay, C. Pichon, and D. Pogosyan, Phys. Rev. D **85**, 023011 (2012).
[18] T. Matsubara, Astrophys. J. Lett. **434**, L43 (1994).
[19] S. Codis, C. Pichon, D. Pogosyan, F. Bernardeau, and T. Matsubara, Mon. Not. R. Astron. Soc. **435**, 531 (2013).
[20] R. Balian and R. Schaeffer, Astron. Astrophys. **220**, 1 (1989).
[21] F. Bernardeau, Astrophys. J. **392**, 1 (1992).
[22] F. Bernardeau, Astron. Astrophys. **291**, 697 (1994).
[23] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, Astrophys. J. Lett. **412**, L9 (1993).
[24] F. Bernardeau, Astrophys. J. **433**, 1 (1994).
[25] C. M. Baugh, E. Gaztanaga, and G. Efstathiou, Mon. Not. R. Astron. Soc. **274**, 1049 (1995).
[26] F. Bernardeau, Astron. Astrophys. **312**, 11 (1996).
[27] F. Bernardeau, Astron. Astrophys. **301**, 309 (1995).
[28] F. Bernardeau and P. Valageas, Astron. Astrophys. **364**, 1 (2000).
[29] P. Valageas, Astron. Astrophys. **382**, 412 (2002).
[30] G. Lavaux and B. D. Wandelt, Astrophys. J. **754**, 109 (2012).
[31] P. M. Sutter, G. Lavaux, B. D. Wandelt, and D. H. Weinberg, Astrophys. J. **761**, 44 (2012).
[32] K. Bolejko and R. A. Sussman, Phys. Lett. B **697**, 265 (2011).
[33] R. van de Weygaert and E. Platen, Int. J. Mod. Phys. Conf. Ser. **01**, 41 (2011).
[34] M. A. Aragon-Calvo, R. van de Weygaert, P. A. Araya-Melo, E. Platen, and A. S. Szalay, Mon. Not. R. Astron. Soc. **404**, L89 (2010).
[35] G. Lavaux and B. D. Wandelt, Mon. Not. R. Astron. Soc. **403**, 1392 (2010).
[36] M. A. Aragon-Calvo and A. S. Szalay, Mon. Not. R. Astron. Soc. **428**, 3409 (2013).
[37] Specifically, we expect perturbation theory to break down later in low density regions; we also expect these regions to probe smaller Lagrangian scales.
[38] J. Binney, N. Dowrick, A. Fisher, and M. Newman, *The Theory of Critical Phenomena* (Oxford University, New York, 1992).
[39] C. Heyde, J. R. Stat. Soc. Ser. B **25**, 392 (1963).
[40] It is to be noted however that it is not necessarily defined for any real values of λ as one expects for instance that the one-point density PDF develops an exponential cutoff, $\mathcal{P}(\rho) \sim \exp(-\lambda_c \rho)$, which implies that $\varphi(\lambda)$ is undefined when $\lambda > \lambda_c$.
[41] This comes from a simple counting of lines: one needs at least $p-1$ lines to connect p points which in turns forces the sum of orders at which each factor is computed to be $2(p-1)$; see [6,21,22,42].
[42] J. N. Fry, Astrophys. J. **279**, 499 (1984).
[43] P. J. E. Peebles, *The Large-Scale Structure of the Universe*, Princeton Series in Physics, edited by S. B. Treiman (Princeton University, Princeton, NJ, 1980).
[44] F. Bernardeau, *Cosmologie: Des Fondements Théoriques aux Observations* (EDP Sciences, Les Ulis, France, 2007).
[45] We could choose $\nu = 3/2$ in order to reproduce the low density asymptotic behavior of the exact solution.
[46] We implicitly assume throughout this paper that the initial conditions are adiabatic so that the perturbations developed out of a single scalar field degree of freedom.
[47] This is actually a superfluous hypothesis but we adopt it for convenience. Transient effects form mildly non-Gaussian initial conditions can for instance be taken into account in this formalism as shown in [48].
[48] P. Valageas, Astron. Astrophys. **382**, 431 (2002).
[49] This relation is *a priori* time dependent but we will omit it in the following.
[50] This inversion is possible as long as $\Psi(\{\tau_k\})$ is a convex function of ρ_k which in practice means before shell crossing.
[51] Thus, if we define $\varphi_{\rm init}(\{\lambda_k\})$ as the cumulant generating function of the density contrasts at initial time, we have $\Psi_{\rm init}(\{\rho_k\}) = \sum_i \rho_i \lambda_i - \varphi_{\rm init}(\{\lambda_k\})$, with $\rho_i = \partial \varphi_{\rm init}/\partial \lambda_i$. It is easy to check that for Gaussian initial conditions, $\varphi_{\rm init}(\{\lambda_i\}) = \sum_i \lambda_i + \frac{1}{2}\sum_{ij}\Sigma_{ij}\lambda_i\lambda_j$, which leads to the expression (25) given previously for Ψ.
[52] F. Bernardeau and R. Schaeffer, Astron. Astrophys. **255**, 1 (1992).
[53] It is beyond the scope of this paper to discuss the validity of this inversion and how it can generally be implemented. Interested readers can find detailed discussions of this construction in [20] in the context of cosmological density PDF and counts in cells statistics, where it has been carried out in particular for the one-point PDF, the properties of which are crucially related to the analytical properties of the cumulant generating function.

[54] In [52] this inversion was initially performed for models of nonlinear clustering (hierarchical models) taking advantage of the general results presented in [20]. The same techniques were later employed in the context of perturbation theory calculations. For all these constructions, the analytical properties of the cumulant generating function follow from the structure of the stationary condition.
[55] Note that in practice π_3 is negative and that only the nonregular parts appearing in the expansion of $\exp(\varphi(\lambda))$ will contribute.
[56] The change of sign in Eq. (45) comes from the path that is *orthogonal* to the real axis.
[57] F. Bernardeau, Astrophys. J. Lett. **390**, L61 (1992).
[58] An alternative approach would be to use predicted amplitude form direct next-to-leading-order perturbation theory calculations. We leave this option for further studies.
[59] F. Bernardeau and L. Kofman, Astrophys. J. **443**, 479 (1995).
[60] The difference can be, to a large extent, interpreted by the nonlinear growth of the reduced skewness.
[61] In practice this is achieved via a series expansion inversion of Eq. (24) then plugged in Eq. (23).
[62] Again when n_s is less than -1 in case of a power law spectrum.
[63] The location of the resulting critical line is only weakly sensitive to the threshold we choose.
[64] It involves expressing P as a function of φ, expressing ρ_2 integrating over ρ_2 and using Cauchy's theorem.
[65] A similar experiment for three-cells of size $R = 11, 12$ and 13 Mpc$/h$ was carried out, producing similar results.
[66] V. Springel, Mon. Not. R. Astron. Soc. **364**, 1105 (2005).
[67] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page *et al.*, Astrophys. J. Suppl. Ser. **192**, 18 (2011).
[68] S. Prunet, C. Pichon, D. Aubert, D. Pogosyan, R. Teyssier, and S. Gottloeber, Astophys. J. Suppl. Ser. **178**, 179 (2008).

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **449,** L105–L109 (2015) doi:10.1093/mnrasl/slv028

The joint statistics of mildly non-linear cosmological densities and slopes in count in cells

Francis Bernardeau,[1,2] Sandrine Codis[1⋆] and Christophe Pichon[1,3]

[1]*CNRS & UPMC, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[2]*CNRS & CEA, UMR 3681, Institut de Physique Théorique, F-91191 Gif-sur-Yvette, France*
[3]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

Accepted 2015 February 10. Received 2015 February 10; in original form 2015 January 15

ABSTRACT

In the context of count-in-cells statistics, the joint probability distribution of the density in two concentric spherical shells is predicted from first principle for sigmas of the order of 1. The agreement with simulation is found to be excellent. This statistics allows us to deduce the conditional one dimensional probability distribution function of the slope within under dense (resp. overdense) regions, or of the density for positive or negative slopes. The former conditional distribution is likely to be more robust in constraining the cosmological parameters as the underlying dynamics is less evolved in such regions. A fiducial dark energy experiment is implemented on such counts derived from Λ cold dark matter simulations.

Key words: methods: numerical – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

With the advent of large galaxy surveys [e.g. the Sloan Digital Sky Survey and in the coming years Euclid (Laureijs et al. 2011), the Large Synoptic Survey Telescope], astronomers have ventured into the era of statistical cosmology and big data. Hence, there is a dire need for them to build tools that can efficiently extract as much information as possible from these huge data sets at high and low redshift. In particular, this means being able to probe the non-linear regime of structure formation. The most commonly used tools to extract statistical information from the observed galaxy distribution are N-point correlation functions (e.g. Scoccimarro et al. 1998) which quantify how galaxies are clustered. In our initially Gaussian Universe, the matter density field is fully described by its two-point correlation function. However departure from Gaussianity occurs when the growth of structure becomes non-linear (at later times or smaller scales), providing information that is not captured by the two-point correlation function but is recorded in part in the three-point correlation function. Obviously N-point correlation functions are increasingly difficult to measure when N increases. They are noisy, subject to cosmic variance and highly sensitive to systematics such as the complex geometry of surveys. It is thus essential to find alternative estimators to extract information from the non-linear regime of structure formation in order to complement these classical probes. This is in particular critical if we are to understand the origin of dark energy, which accounts for ∼70 per cent of the energy budget of our Universe.

One such method to accurately probe the non-linear regime is to implement perturbation theory in a highly symmetric configuration (spherical or cylindrical symmetry) for which the full joint cumulant generating functions can be constructed. Such constructions take advantage of the fact that non-linear solutions to the gravitational dynamical equations (the so-called spherical collapse model) are known exactly. Corresponding observables, such as galaxy counts in concentric spheres or discs, then yield very accurate analytical predictions in the mildly non-linear regime, well beyond what is usually achievable using other estimators. The corresponding symmetry implies that the most likely dynamical evolution (amongst all possible mapping between the initial and final density field) is that corresponding to the spherical collapse for which we can write an explicit linear to non-linear mapping. This has been demonstrated in the limit of zero variance using direct diagram resummations (Bernardeau 1992, 1994)[1] which was later shown to correspond to a saddle approximation (Juszkiewicz, Bouchet & Colombi 1993; Valageas 2002; Bernardeau, Pichon & Codis 2014). The key point on which this whole paper is based upon, is that the zero variance limit is shown to provide a remarkably good working model for finite variances (Bernardeau 1994; Bernardeau et al. 2014).

This formalism also allows us to weigh non-uniformly different regions of the universe making possible to take into account the fact that the noise structure in surveys is not homogenous. For instance, low-density regions are probed by fewer galaxies. Conversely, on dynamical grounds, we also expect the level of non-linearity in

[1] The original derivations were actually derived from the hierarchical model that aimed at describing the fully non-linear regime (Balian & Schaeffer 1989).

⋆ E-mail: codis@iap.fr

the field to be inhomogenous: low-density regions are less non-linear. Hence, it is of interest to build statistical estimators which probe the mildly non-linear regime and that can be tuned to probe subsets of the field, offering the best compromise between these constraints. In the context of the cosmic density field, the construction of conditional distributions naturally leads to the elaboration of joint probability distribution functions (PDF hereafter) of the density in concentric cells.

Following Bernardeau, Pichon & Codis 2014 (hereafter BPC), we propose in Section 2 to extend one-point statistics of density profiles and to the full joint PDF of the density in two concentric spheres of different radii. This is obtained using perturbation theory core results on the cumulant generating function, the double inverse Laplace transform of which is then computed from brute force numerical integration. From that PDF, we will also present the statistics of density profiles restricted to underdense (resp. overdense) regions, and the statistics of density restricted to positive (resp. negative) slopes (Section 3). Theoretical predictions will be shown to be in very good agreement with simulations in the mildly non-linear regime. Dependence with redshift will also be discussed. Finally, Section 4 presents a simple fiducial dark energy experiment, while Section 5 wraps up.

2 THE TWO-CELL DENSITY STATISTICS

For the sake of clarity, let us present and briefly comment the formalism. We consider two spheres $\mathcal{S}_i$ of radius R_i ($i = 1, 2$) centred on a given location of space $\boldsymbol{x}_0$. Our goal is to derive the joint PDF of the density in $\mathcal{S}_1$ and $\mathcal{S}_2$ denoted $\hat{\rho}_i$ and rescaled so that $\langle\hat{\rho}_i\rangle = 1$.

2.1 The cumulant generating function

In the cases we are interested in, the joint statistical properties of $\hat{\rho}_1$ and $\hat{\rho}_2$ are fully encoded in their moment generating function

$$\mathcal{M}_{R_1R_2}(\lambda_1,\lambda_2) = \sum_{p,q=0}^{\infty} \langle\hat{\rho}_1^p\hat{\rho}_2^q\rangle \frac{\lambda_1^p\lambda_2^q}{p!\,q!}\,, \tag{1}$$

$$= \langle\exp(\lambda_1\hat{\rho}_1 + \lambda_2\hat{\rho}_2)\rangle\,, \tag{2}$$

that can be related to the *cumulant* generating function, $\varphi_{R_1R_2}(\lambda_1,\lambda_2)$, through $\mathcal{M}_{R_1R_2}(\lambda_1,\lambda_2) = \exp\left[\varphi_{R_1R_2}(\lambda_1,\lambda_2)\right]$, so that

$$\exp\left[\varphi_{R_1R_2}\right] = \int \mathrm{d}\hat{\rho}_1\mathrm{d}\hat{\rho}_2\mathcal{P}_{R_1R_2}\exp(\lambda_1\hat{\rho}_1 + \lambda_2\hat{\rho}_2), \tag{3}$$

where $\mathcal{P}_{R_1R_2}(\hat{\rho}_1,\hat{\rho}_2)$ is the joint PDF of having density $\hat{\rho}_1$ in $\mathcal{S}_1$ and $\hat{\rho}_2$ in $\mathcal{S}_2$. We will now exploit a theoretical construction that permits the explicit calculation of $\mathcal{P}_{R_1R_2}(\hat{\rho}_1,\hat{\rho}_2)$.

2.1.1 Upshot

As we will sketch in the following, this theoretical construction yields the explicit time dependence of the *Legendre transform* of $\varphi_{R_1R_2}(\lambda_1,\lambda_2)$ in the quasi-linear regime. Such a Legendre transform is defined as

$$\Psi_{R_1R_2}(\rho_1,\rho_2) = \lambda_1\rho_1 + \lambda_2\rho_2 - \varphi_{R_1R_2}(\lambda_1,\lambda_2), \tag{4}$$

where ρ_i are determined implicitly by the stationary conditions

$$\lambda_i = \frac{\partial}{\partial\rho_i}\Psi_{R_1R_2}(\rho_1,\rho_2)\,,\quad i = 1,2\,. \tag{5}$$

The fundamental relation is then that, in the limit of zero variance, this Legendre transforms taken at two different times, $\Psi(\rho_1,\rho_2;\eta)$ and $\Psi'(\rho_1,\rho_2;\eta')$, take the *same* value

$$\Psi_{R_1R_2}(\rho_1,\rho_2;\eta) = \Psi_{R_1'R_2'}(\rho_1',\rho_2';\eta')\,, \tag{6}$$

provided that $\rho_i R_i^3 = \rho_i' R_i'^3$, and that ρ_i' and ρ_i are linked together through the non-linear dynamics of spherical collapse. The origin of this equation will be sketched in the following Section 2.1.2.

Equation (6), when applied to an arbitrarily early time η', yields a relation between $\Psi(\rho_1,\rho_2;\eta)$ and the statistical properties of the *initial* density fluctuations. In particular, for Gaussian initial conditions, $\Psi(\rho_1,\rho_2;\eta_i)$ can easily be calculated and expressed in terms of elements of covariance matrices,

$$\Psi_{R_1R_2}(\rho_1,\rho_2;\eta_i) = \frac{1}{2}\sum_{i,j\le2}\Xi_{ij}(\rho_i-1)(\rho_j-1)\,, \tag{7}$$

where Ξ_{ij} is the *inverse* of the matrix of covariances, $\Sigma_{ij} = \langle\tau_i\tau_j\rangle$, between the initial density contrasts τ_i in the two concentric spheres of radii R_i. One can then write the cumulant generating function at any time through the spherical collapse mapping between one final density at time η in a sphere of radius R_i and one initial contrast in a sphere centred on the same point and with radius $R_i' = R_i\rho_i^{1/3}$ (so as to encompass the same total mass); it can be written formally as

$$\rho_i = \zeta_{\rm SC}(\eta,\tau_i) \approx \frac{1}{(1-D_+(\eta)\tau/\nu)^\nu}\,, \tag{8}$$

where, for the sake of simplicity, we use here a simple prescription, with $D_+(\eta)$ the linear growth factor and $\nu = 21/13$ to reproduce the high-z skewness.

Recall that only $\Psi_{R_1R_2}(\rho_1,\rho_2)$ is easily computed. The statistically relevant cumulant generating function, $\varphi_{R_1R_2}(\lambda_1,\lambda_2)$, is only accessible via equation (4) through an inverse Legendre transform which brings its own complications. In particular, note that all values of λ_i are not accessible due to the fact that the ρ_i–λ_i relation cannot always be inverted. This is signalled by the fact that the determinant of the transformation vanishes, e.g. $\mathrm{Det}\left[\partial\rho_i\partial\rho_j\Psi(\rho_1,\rho_2)\right] = 0$. This condition is met both for finite values of ρ_i and λ_i. The corresponding contour lines of $\varphi(\lambda_1,\lambda_2)$ was investigated in BPC and successfully compared to simulation.

2.1.2 Motivation

It is beyond the scope of this letter to re-derive equations (4)–(6) – a somewhat detailed presentation can be found in Valageas (2002) and in BPC – but we can give a hint of where it comes from: it is always possible to express any ensemble average in terms of the statistical properties of the initial density field so that we can formally write

$$\exp\left[\varphi\right] = \int\mathcal{D}\tau_1\mathcal{D}\tau_2\,\mathcal{P}(\tau_1,\tau_2)\exp\left(\lambda_1\rho_1(\tau_1) + \lambda_2\rho_2(\tau_2)\right)\,. \tag{9}$$

As the present-time densities ρ_i can arise from different initial contrasts, the above-written integration is therefore a path integral (over all the possible paths from initial conditions to present-time configuration) with measure $\mathcal{D}\tau_1\mathcal{D}\tau_2$ and known initial statistics $\mathcal{P}(\tau_1,\tau_2)$. Let us assume here that the initial PDF is Gaussian so that

$$\mathcal{P}(\tau_1,\tau_2)\mathrm{d}\tau_1\mathrm{d}\tau_2 = \frac{\sqrt{\det\Xi}\exp\left[-\Psi(\tau_1,\tau_2)\right]}{2\pi}\mathrm{d}\tau_1\mathrm{d}\tau_2\,, \tag{10}$$

with Ψ then a quadratic form.

In the regime where the variance of the density field is small, equation (9) is dominated by the path corresponding to the most likely configurations. As the constraint is spherically symmetric, this most likely path should also respect spherical symmetry. It is therefore bound to obey the spherical collapse dynamics. Within this regime equation (9) becomes

$$\exp[\varphi] \simeq \int \mathrm{d}\tau_1 \mathrm{d}\tau_2\, \mathcal{P}(\tau_1, \tau_2) \exp\left(\lambda_1 \zeta_{\rm SC}(\tau_1) + \lambda_2 \zeta_{\rm SC}(\tau_2)\right), \qquad (11)$$

where the most likely path, $\rho_i = \zeta_{\rm SC}(\eta, \tau_i)$ is the one-to-one spherical collapse mapping between one final density at time η and one initial density contrast as already described. The integration on the r.h.s. of equation (11) can now be carried by using a steepest descent method, approximating the integral as its most likely value, where $\lambda_1\rho_1(\tau_1) + \lambda_2\rho_2(\tau_2) - \Psi(\tau_1, \tau_2)$ is stationary. It eventually leads to the fundamental relation (6) when its right-hand side is computed at initial time (and the fact that (6) is valid for any times η and η' is obtained when the same reasoning is applied twice, for the two different times).

The purpose of this letter is to confront numerically computations of the two-cell PDF derived from the expression of $\varphi(\lambda_1, \lambda_2)$ with measurements in numerical simulations.

2.2 The two-cell PDF using inverse Laplace transform

Once the cumulant generating function is known in equation (3), the two-cell PDF, $\mathcal{P}(\hat{\rho}_1, \hat{\rho}_2)$, is obtained by a 2D inverse Laplace transform of $\varphi(\lambda_1, \lambda_2)$

$$\mathcal{P} = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda_1}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda_2}{2\pi\mathrm{i}} \exp\left(-\sum_{i=1,2} \hat{\rho}_i\lambda_i + \varphi(\lambda_1, \lambda_2)\right), \qquad (12)$$

with φ given by equations (4)–(6). From this equation, it is straightforward to deduce the joint PDF, $\hat{\mathcal{P}}(\hat{\rho}, \hat{s})$, for the density, $\hat{\rho} = \hat{\rho}_1$ and the slope $\hat{s} \equiv (\hat{\rho}_2 - \hat{\rho}_1)R_1/\Delta R$, ΔR being $R_2 - R_1$, as

$$\hat{\mathcal{P}} = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\lambda}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\mu}{2\pi\mathrm{i}} \exp\left(-\hat{\rho}\lambda - \hat{s}\mu + \varphi(\lambda, \mu)\right), \qquad (13)$$

with $\lambda = \lambda_1 + \lambda_2$, $\mu = \lambda_2\Delta R/R_1$. Following this definition, $\varphi(\lambda, \mu)$ is also the Legendre transform of $\Psi(\hat{\rho}_1, \hat{s} = (\hat{\rho}_2 - \hat{\rho}_1)\, R_1/\Delta R)$.

In order to numerically compute equation (12), we simply choose the imaginary path $(\lambda_1, \lambda_2) = \mathrm{i}(n_1\Delta\lambda, n_2\Delta\lambda)$ where n_1 and n_2 are (positive or negative) integers and the step $\Delta\lambda$ has been set to 0.15. The maximum value of λ_i used here is 75 resulting into a discretization of the integrand on 1000^2 points. Fig. 1 compares the result of the numerical integration of equation (12) to simulations. The corresponding dark matter simulation (carried out with GADGET2 Springel 2005) is characterized by the following Λ cold dark matter (ΛCDM) cosmology: $\Omega_{\rm m} = 0.265$, $\Omega_\Lambda = 0.735$, $n = 0.958$, $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\sigma_8 = 0.8$, $\Omega_{\rm b} = 0.045$ within one standard deviation of *Wilkinson Microwave Anisotropy Probe* 7 results (Komatsu et al. 2011). The box size is 500 Mpc h^{-1} sampled with 1024^3 particles, the softening length 24 kpc h^{-1}. Initial conditions are generated using MPGRAFIC (Prunet et al. 2008). An Octree is built to count efficiently all particles within concentric spheres of radii between $R = 10$ and 11Mpc h^{-1}. The centre of these spheres is sampled regularly on a grid of 10 Mpc h^{-1} aside, leading to 117 649 estimates of the density per snapshot. Note that the cells overlap for radii larger than 10 Mpc h^{-1}.

The convergence of our numerical scheme is investigated by varying the number of points. Fig. 2 shows that the numerical integration of the slope PDF has reached 1 per cent precision for the displayed range of slopes. Obviously, the integration is very precise for low values of the slope and requires a largest number of points for the large-slope tails.

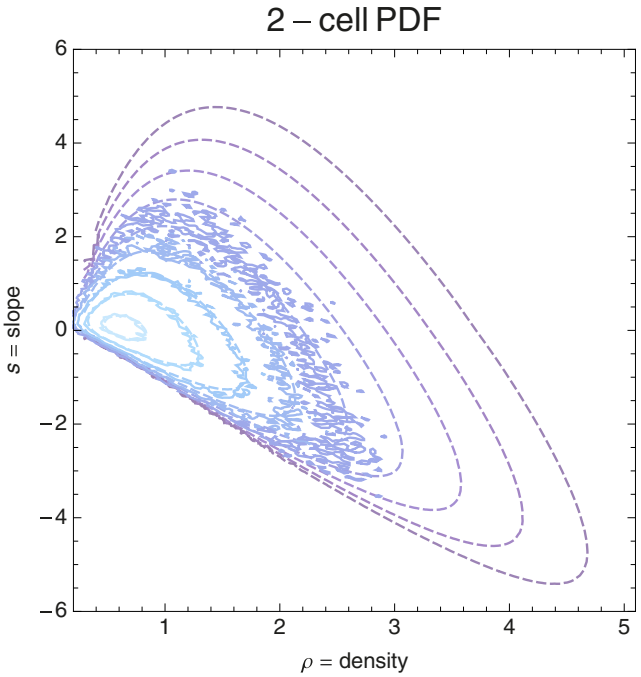


Figure 1. Joint PDF of the slope (s) and the density (ρ) as given by equation (13) for two concentric spheres of radii $R_1 = 10$ Mpc h^{-1} and $R_2 = 11$ Mpc h^{-1} at redshift $z = 0.97$. Dashed contours corresponds to log $\mathcal{P} = 0, -1/2, -1, \ldots, -3$ for the theory. The corresponding measurements are shown as a solid line.

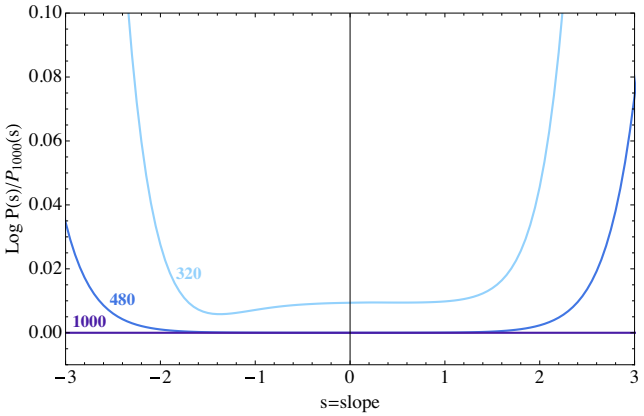


Figure 2. Dependence of the PDF of the slope on the number of points used in the numerical integration in (12). The reference PDF is computed using 1000^2 points (dark blue) and is compared to the result of the numerical integration when using 320^2 (blue) and 480^2 (light blue) points.

3 CONDITIONAL DISTRIBUTIONS

3.1 Slope in subregions

Once the full two-cell PDF is known, it is straightforward to derive predictions for density profiles restricted to underdense

$$\mathcal{P}(\hat{s}|\hat{\rho} < 1) = \frac{\int_0^1 \mathrm{d}\hat{\rho}\; \hat{\mathcal{P}}(\hat{\rho}, \hat{s})}{\int_{-\infty}^{\infty} \mathrm{d}\hat{s} \int_0^1 \mathrm{d}\hat{\rho}\; \hat{\mathcal{P}}(\hat{\rho}, \hat{s})}, \qquad (14)$$

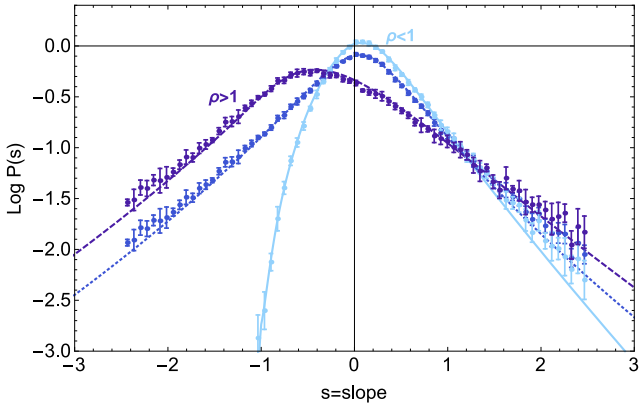


Figure 3. Density profiles in underdense (solid light blue), overdense (dashed purple) and all regions (dashed blue) for cells of radii $R_1 = 10$ Mpc h^{-1} and $R_2 = 11$ Mpc h^{-1} at redshift $z = 0.97$. Predictions are successfully compared to measurements in simulations (points with error bars).

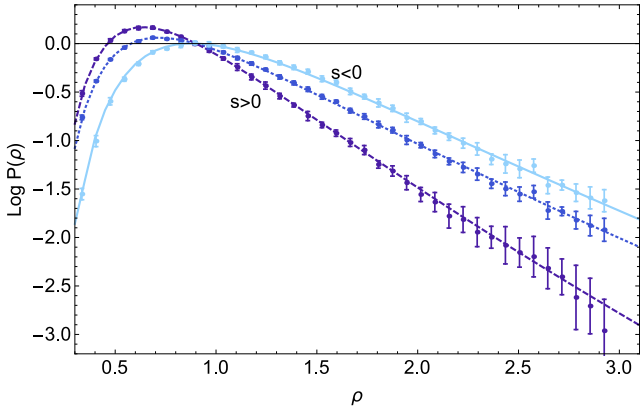


Figure 4. Density PDF in negative slope (solid light blue), positive slope (dashed purple) and all regions (dashed blue) for cells of radii $R_1 = 10$ Mpc h^{-1} and $R_2 = 11$ Mpc h^{-1} at redshift $z = 0.97$. Predictions are successfully compared to measurements in simulations (points with error bars).

and overdense regions

$$\mathcal{P}(\hat{s}|\hat{\rho} > 1) = \frac{\int_1^\infty \mathrm{d}\hat{\rho}\ \hat{\mathcal{P}}(\hat{\rho}, \hat{s})}{\int_{-\infty}^{\infty} \mathrm{d}\hat{s} \int_0^1 \mathrm{d}\hat{\rho}\ \hat{\mathcal{P}}(\hat{\rho}, \hat{s})}\,. \tag{15}$$

Fig. 3 displays these predicted density profiles in underdense and overdense regions compared to the measurements in our simulation. A very good agreement is found with some slight departures in the large-slope tail of the distribution. As expected, the underdense slope PDF peaks towards positive slope, while the overdense PDF peaks towards negative slope. The constrained negative tails are more sensitive to the underlying constraint, providing improved leverage for measuring the underlying cosmological parameters.

3.2 Density in regions of given slope

Conversely, one can study the statistics of the density given constraints on the slope. For instance, the density PDF in regions of negative slope reads

$$\mathcal{P}(\hat{\rho}|\hat{s} < 0) = \frac{\int_{-\infty}^0 \mathrm{d}\hat{s}\ \hat{\mathcal{P}}(\hat{\rho}, \hat{s})}{\int_0^\infty \mathrm{d}\hat{\rho} \int_{-\infty}^0 \mathrm{d}\hat{s}\ \hat{\mathcal{P}}(\hat{\rho}, \hat{s})}\,. \tag{16}$$

Fig. 4 displays the predicted density PDF in regions of positive or negative slope. As expected, the density is higher in regions of negative slope. An excellent agreement with simulations is found.

3.3 Redshift evolution

Fig. 5 displays the density profiles in underdense and overdense regions as measured in the simulation for a range of redshifts. This figure shows that the high-density subset for moderately negative slopes is particularly sensitive to redshift evolution, which suggests that dark energy investigations should focus on such range of slopes and regions.

4 FIDUCIAL DARK ENERGY EXPERIMENT

Let us conduct the following fiducial experiment. Consider a set of 10 000 concentric spheres, and measure for each pair the slope and the density, $\{\hat{\rho}_i, \hat{s}_i\}$. Recall that the cosmology is encoded in the parametrization of the spherical collapse on the

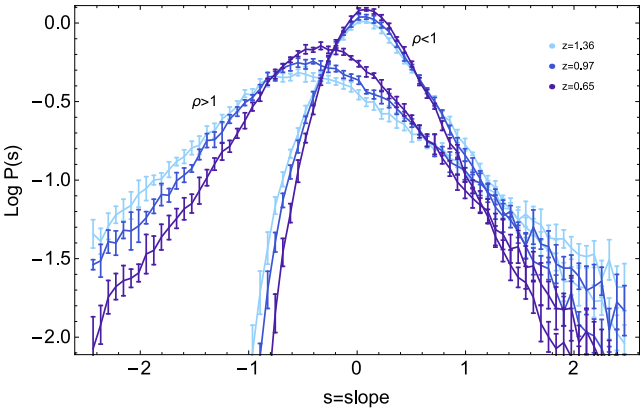


Figure 5. Same as Fig. 3 for a range of redshifts as labelled. Only the underdense ($\rho < 1$) and the overdense ($\rho > 1$) PDFs are shown.

one hand (ν),[2] and on the linear power spectrum, $P_k^{\rm lin}$, (via the covariance matrix, $\Sigma_{ij} = \int P_k^{\rm lin}(k) W(R_i k) W(R_j k) d^3k/(2\pi)^3$ with $W(k) = 3(\sin(k)/k - \cos(k))/k^2$) on the other hand. For scale invariant power spectra with power index n, given equation (8), we have a three parameter (n, ν, σ) set of models where σ^2 is the variance of the inner densities $\sigma^2(z) = \rho_{\rm c}(z)^2 \left\langle \hat{\rho}_1^2 \right\rangle$, $\rho_{\rm c}$ being the critical density of the Universe at redshift z. For a parametrized PDF, $P_{n,\nu,\sigma}(\hat{\rho}, \hat{s})$ given by equation (13), we can compute the log-likelihood of the set as $\mathcal{L}(n, \nu, \sigma) = \sum_i \log P_{n,\nu,\sigma}(\hat{\rho}_i, \hat{s}_i)$. Fig. 6 displays the corresponding likelihood contours at 1σ, 3σ, and 5σ in the simple case in which only one parameter (σ here) varies. This experiment mimics the precision expected from a survey of useful volume of about $(350h^{-1}\mathrm{Mpc})^3$ which is found to be at the percent level. This work improves the findings of BPC which relies on a low-density approximation for the joint PDF.

5 CONCLUSION

Extending the analysis of BPC, predictions for the *joint* PDF of the density within two concentric spheres was straightforwardly

[2] The dependence of the spherical collapse on cosmology is at the percent level as discussed in Bernardeau et al. (2002).

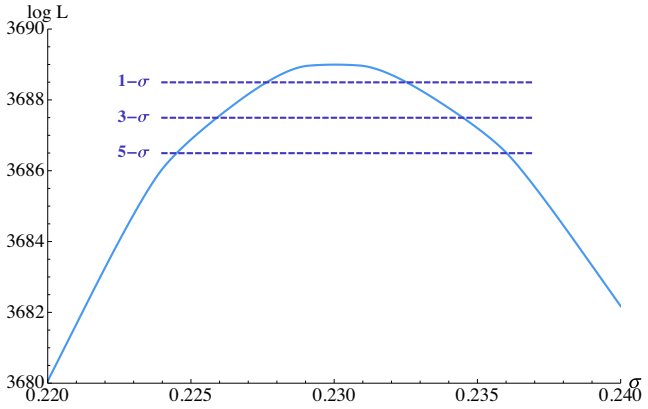


Figure 6. Log-likelihood for a fiducial experiment involving 10 000 concentric spheres of 10 and 11 Mpc h^{-1} measured in our simulation. The model here only depends on the variance σ (ν and n are fixed). The contours at 1σ, 3σ, and 5σ centred on the true value 0.23 are displayed with dark blue dashed lines. The same experiment can be carried out when the three parameters vary.

implemented for a given cosmology as a function of redshift in the mildly non-linear regime. The agreement with measurements in simulation was shown on Figs 1, 3, and 4 to be very good, including in the quasi-linear regime where standard perturbation theory normally fails. A fiducial dark energy experiment was implemented on counts derived from ΛCDM simulations and was illustrated on Fig. 6.

Such statistics will prove useful in upcoming surveys as they allow us to probe differentially the slope of the density in regions of low or high density. It can serve as a statistical indicator to test gravity and dark energy models and/or probe key cosmological parameters in carefully chosen subsets of surveys. The theory of count in cells could be applied to 2D cosmic shear maps so as to predict the statistics of projected density profiles. Velocity profiles and combined probes involving the density and velocity fields should also be within reach of this formalism.

ACKNOWLEDGEMENTS

This work is partially supported by the grants ANR-12-BS05-0002 and ANR-13-BS05-0005 of the French *Agence Nationale de la Recherche*. The simulations were run on the `Horizon` cluster. We acknowledge support from S. Rouberol for running the cluster for us.

REFERENCES

Balian R., Schaeffer R., 1989, A&A, 220, 1
Bernardeau F., 1992, ApJ, 390, L61
Bernardeau F., 1994, A&A, 291, 697
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bernardeau F., Pichon C., Codis S., 2014, Phys. Rev. D, 90, 103519 (BPC)
Juszkiewicz R., Bouchet F. R., Colombi S., 1993, ApJ, 412, L9
Komatsu E. et al., 2011, ApJS, 192, 18
Laureijs R. et al., 2011, preprint (arXiv:e-prints)
Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R., Gottloeber S., 2008, ApJS, 178, 179
Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, ApJ, 496, 586
Springel V., 2005, MNRAS, 364, 1105
Valageas P., 2002, A&A, 382, 412

This paper has been typeset from a TEX/LATEX file prepared by the author.

Conclusion

During my thesis, my interest has been the theoretical understanding of the large-scale structure of the Universe, in particular addressing some of the challenges that the field of large-scale structure studies needs to overcome with the advent of large galaxy surveys like Euclid, namely redshift space distortion, intrinsic alignments, non-linear clustering. In this last section, I summarize the main results I obtained, present concluding remarks and suggest possible extensions.

Galaxy morphology within the cosmic web

I first investigated how the morphology of dark halos and galaxies are correlated with the cosmic web from a numerical and theoretical point of view.

Spin of simulated galaxies I measured in simulations how the spin of galaxies is correlated to the direction of the filament they are embedded in and found that these correlations were mass-dependent ([Codis et al., 2012, Dubois et al., 2014]). This signal can be qualitatively understood in the context of hierarchical structure formation, where small galaxies form first and merge together to form larger ones. Indeed, the first generation of small galaxies form when the walls wind up to give rise to the filaments and during this process protogalaxies acquire spin mostly aligned with the filaments. When they later merge inside filaments by catching each other up during their race towards the nodes, they naturally form bigger-size objects with spin perpendicular to the filaments. It has to be noted that recent observations (in particular on the SDSS) have reported some hints of a correlation between galaxy morphology and the cosmic web as predicted by dark matter and hydrodynamical simulations.

Spin acquisition by anisotropic tidal torquing In this manuscript, I also presented a detailed analytic prediction ([Codis et al., 2015b]) which explicitly takes into account the anisotropy of the cosmic web to reproduce from first principles the correlation between galactic spins and filaments as a function of halo mass. The idea was to develop an extension of the theory of linear tidal torquing where the cosmic web was made explicit by requiring that galaxies do not form everywhere but only in the filaments. The outcome of such a theory is a mass-dependent orientation of the spin in agreement with what has been measured in simulations. The typical quadrants of opposite vorticity seen in simulations ([Laigle et al., 2015]) are also nicely explained within the context of this Anisotropic Tidal Torque Theory.

Intrinsic alignments The impact of the cosmic web on galaxies is of prime importance in order to understand not only how the Hubble sequence sets up but also to investigate a possible important contamination of weak lensing surveys : intrinsic alignments. I recently measured

this contamination directly from the 160,000 galaxies of a state-of-the-art hydrodynamical simulation that allowed me to take into account non-linear gas dynamics, baryonic physics and the anisotropy of the cosmic web ([Codis et al., 2015a, Chisari et al., 2015]). A distinctive mass and colour-dependence has been measured and illustrates the need to use cosmological hydrodynamical simulations to tackle intrinsic alignment effects that could be a concern for future missions like Euclid.

Cosmological probes

I also studied how the statistics of the large-scale structure can be used to probe the ΛCDM model – and its extensions – by means of perturbation theory, topology and galaxy number counts in the weakly and mildly non-linear regime.

Perturbation theory The dynamical equations that govern the evolution of the density field in the Universe are highly non-linear. They can be analytically solved only in the configuration of maximal symmetry which is an initially spherically symmetric perturbation, corresponding to the so-called spherical collapse. In any other regime, there is no known solution and one has to rely on perturbation theory on large scales or simulations. Standard perturbation theory is valid on very large scales but breaks down quickly. Resummation schemes have then been implemented to achieve better convergence and indeed succeeded in extending the regime of validity for analytical predictions. In [Taruya et al., 2012], we computed the density power spectrum at next-to-leading order within the Regularised Perturbation Theory, one extension of the standard approach. An agreement with simulations was found at one-percent level in the weakly non-linear regime; we produced a fast public code that allows one to quickly compute the density power spectrum for a wide range of cosmological models.

Topology Another approach is to consider the topological features of the density field which represent a complementary cosmological probe containing robust information. The topology of a 3D field can be described by only four functionals – named after Minkowski – that can be analytically computed when non-Gaussianities are weak enough. In [Codis et al., 2013], I developed the theory of these estimators to all order in non-Gaussianity. I also showed how redshift space distortion can be accounted for within this context. The formalism used involves an expansion of the joint probability density function of the field and its derivatives to all orders in non-Gaussianity by means of a Gram-Charlier expansion, a rewriting of the equations in terms of invariant quantities (invariant under 3D rotations for real space and 2D rotations on the sky for redshift space). With upcoming 3D spectroscopic surveys such as Euclid, the statistical analysis of the topology of our redshift-distorted Universe will allow us to robustly measure weighted moments of the multi-spectra as a function of redshift, and henceforth quantify the cosmic evolution of the equation of state of dark energy and possible departures from General Relativity. The challenge here is in the lower level of symmetry compared to real space (no longer isotropic) that complexifies the formalism by adding more degrees of freedom. I specifically investigated how topological estimators in redshift space can be used to estimate the cosmic evolution of the growth of structure (the so-called β parameter).

Density profiles Number counts of galaxies in concentric spheres provide another complementary observable. The joint statistics of these mean concentric densities can be predicted from

perturbation theory with impressive accuracy even in the mildly non-linear regime as spherical symmetry and Gauss' theorem allow one to map initial densities to late-time configurations in concentric spheres. This unique formalism can be applied to various numbers of spheres. In [Bernardeau et al., 2014], we predicted the profiles (as the density difference between two concentric spheres) in the quasi-linear regime, and studied the effect of modified gravity models on these counts through a parametrisation of the spherical collapse. In [Bernardeau et al., 2015], we showed how to compute the conditional density profiles within underdense or overdense subregions.

Prospects

During my thesis I analysed cosmological simulations – both dark matter only and hydrodynamical simulations – in the context of galaxy formation. The motivation was to understand how galaxies are shaped by their environment. I also worked on the theory of gravitational clustering through perturbation theory and its extensions, the topology and geometry of the large-scale structure and the theory of protogalactic spin acquisition. Interesting prospects include

- galaxy counts and how the theory of count-in-cells could be applied to 2D cosmic shear maps so as to predict the statistics of projected density profiles in the context of weak lensing but also to velocity profiles and cosmic variance among others;
- cosmological probes and the finding of optimal observables that can extract non-linear cosmological information from the large-scale structure. One promising way is to understand which (combination of) estimators is best, including topology and galaxy counts and in particular compare their sensitivity to different models of modified gravity and different equations of state of dark energy;
- the geometry of the cosmic web. It would be worth developing an Alcock-Paczynski test based on the length of the filaments. Another interesting analysis could be a theoretical study of the walls (which is related to the numerical study of voids that is becoming a burning issue nowadays) and of the connectivity of the filaments. This latter point is also important to understand how many streams feed galaxies which is one of the cornerstones of the understanding of galaxy formation;
- intrinsic alignments (e.g in the perspective of Euclid) and how numerical analysis can be used to define an "optimised strategy" for cosmic shear experiments so as to reduce and leverage the effect of intrinsic alignments by specific geometrical choices and/or population selections (on the different properties of galaxies such as their colour).

Bibliography

[Aarseth et al., 1979] Aarseth, S. J., Turner, E. L., and Gott, III, J. R. (1979). N-body simulations of galaxy clustering. I - Initial conditions and galaxy collapse times. *ApJ*, 228:664–683.

[Abazajian et al., 2003] Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Anderson, S. F., Annis, J., Bahcall, N. A., Baldry, I. K., Bastian, S., Berlind, A., and et al. (2003). The First Data Release of the Sloan Digital Sky Survey. *AJ*, 126:2081–2086.

[Adler, 1981] Adler, R. J. (1981). *The Geometry of Random Fields*. Chichester: Wiley.

[Alimi et al., 2012] Alimi, J.-M., Bouillot, V., Rasera, Y., Reverdy, V., Corasaniti, P.-S., Balmes, I., Requena, S., Delaruelle, X., and Richet, J.-N. (2012). DEUS Full Observable {\Lambda}CDM Universe Simulation: the numerical challenge. *ArXiv e-prints*.

[Amendola, 1996] Amendola, L. (1996). Non-Gaussian likelihood function and COBE data. *MNRAS*, 283:983–989.

[Angulo et al., 2012] Angulo, R. E., Springel, V., White, S. D. M., Jenkins, A., Baugh, C. M., and Frenk, C. S. (2012). Scaling relations for galaxy clusters in the Millennium-XXL simulation. *MNRAS*, 426:2046–2062.

[Appel, 1985] Appel, A. W. (1985). An Efficient Program for Many-Body Simulation. *SIAM Journal on Scientific and Statistical Computing, vol. 6, no. 1, January 1985, p. 85-103.*, 6:85–103.

[Aragón-Calvo et al., 2007a] Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., and van der Hulst, J. M. (2007a). The multiscale morphology filter: identifying and extracting spatial patterns in the galaxy distribution. *A&A*, 474:315–338.

[Aragón-Calvo et al., 2010] Aragón-Calvo, M. A., Platen, E., van de Weygaert, R., and Szalay, A. S. (2010). The Spine of the Cosmic Web. *ApJ*, 723:364–382.

[Aragón-Calvo et al., 2007b] Aragón-Calvo, M. A., van de Weygaert, R., Jones, B. J. T., and van der Hulst, J. M. (2007b). Spin Alignment of Dark Matter Halos in Filaments and Walls. *ApJ Let.*, 655:L5–L8.

[Aragon-Calvo and Yang, 2014] Aragon-Calvo, M. A. and Yang, L. F. (2014). The hierarchical nature of the spin alignment of dark matter haloes in filaments. *MNRAS*, 440:L46–L50.

[Aubert et al., 2004] Aubert, D., Pichon, C., and Colombi, S. (2004). The origin and implications of dark matter anisotropic cosmic infall on =L haloes. *MNRAS*, 352(2):376–398.

[Bailin and Steinmetz, 2005] Bailin, J. and Steinmetz, M. (2005). Internal and External Alignment of the Shapes and Angular Momenta of ΛCDM Halos. *ApJ*, 627:647–665.

[Bardeen et al., 1986] Bardeen, J. M., Bond, J. R., Kaiser, N., and Szalay, A. S. (1986). The statistics of peaks of gaussian random fields. *ApJ*, 304:15–61.

[Barnes and Efstathiou, 1987] Barnes, J. and Efstathiou, G. (1987). Angular momentum from tidal torques. *ApJ*, 319:575–600.

[Barnes and Hut, 1986] Barnes, J. and Hut, P. (1986). A hierarchical O(N log N) force-calculation algorithm. *Nature*, 324:446–449.

[Barrow et al., 1985] Barrow, J. D., Bhavsar, S. P., and Sonoda, D. H. (1985). Minimal spanning trees, filaments and galaxy clustering. *MNRAS*, 216:17–35.

[Bartelmann and Schneider, 2001] Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. *Phys. Rep.*, 340:291–472.

[Baugh, 2006] Baugh, C. M. (2006). A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports on Progress in Physics*, 69:3101–3156.

[Berlind and Weinberg, 2002] Berlind, A. A. and Weinberg, D. H. (2002). The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *ApJ*, 575:587–616.

[Bernardeau, 1994a] Bernardeau, F. (1994a). Skewness and kurtosis in large-scale cosmic fields. *ApJ*, 433:1–18.

[Bernardeau, 1994b] Bernardeau, F. (1994b). The effects of smoothing on the statistical properties of large-scale cosmic fields. *A&A*, 291:697–712.

[Bernardeau, 1995] Bernardeau, F. (1995). The angular correlation hierarchy in the quasilinear regime. *A&A*, 301:309–+.

[Bernardeau et al., 2015] Bernardeau, F., Codis, S., and Pichon, C. (2015). The joint statistics of mildly non-linear cosmological densities and slopes in count in cells. *MNRAS*, 449:L105–L109.

[Bernardeau et al., 2002] Bernardeau, F., Colombi, S., Gaztañaga, E., and Scoccimarro, R. (2002). Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rep.*, 367:1–248.

[Bernardeau et al., 2008] Bernardeau, F., Crocce, M., and Scoccimarro, R. (2008). Multipoint propagators in cosmological gravitational instability. *Phys. Rev. D*, 78(10):103521.

[Bernardeau et al., 2010] Bernardeau, F., Crocce, M., and Sefusatti, E. (2010). Multipoint propagators for non-Gaussian initial conditions. *Phys. Rev. D*, 82(8):083507.

[Bernardeau and Kofman, 1995] Bernardeau, F. and Kofman, L. (1995). Properties of the cosmological density distribution function. *ApJ*, 443:479–498.

[Bernardeau et al., 2014] Bernardeau, F., Pichon, C., and Codis, S. (2014). Statistics of cosmic density profiles from perturbation theory. *Phys. Rev. D*, 90(10):103519.

[Bernstein and Norberg, 2002] Bernstein, G. M. and Norberg, P. (2002). Shape Alignments of Satellite Galaxies. *AJ*, 124:733–736.

[Blake et al., 2011a] Blake, C., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S., Davis, T., Drinkwater, M. J., Forster, K., Gilbank, D., and et al. (2011a). The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9. *MNRAS*, 415:2876–2891.

[Blake et al., 2011b] Blake, C., Kazin, E. A., Beutler, F., Davis, T. M., Parkinson, D., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S., and et al. (2011b). The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations. *MNRAS*, 418:1707–1724.

[Blanton et al., 2005] Blanton, M. R., Eisenstein, D., Hogg, D. W., Schlegel, D. J., and Brinkmann, J. (2005). Relationship between Environment and the Broadband Optical Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 629:143–157.

[Blazek et al., 2012] Blazek, J., Mandelbaum, R., Seljak, U., and Nakajima, R. (2012). Separating intrinsic alignment and galaxy-galaxy lensing. *JCAP*, 5:41.

[Blazek et al., 2011] Blazek, J., McQuinn, M., and Seljak, U. (2011). Testing the tidal alignment model of galaxy intrinsic alignment. *JCAP* , 5:10.

[Blazek et al., 2015] Blazek, J., Vlah, Z., and Seljak, U. (2015). Tidal alignment of galaxies. *ArXiv e-prints*.

[Blinnikov and Moessner, 1998] Blinnikov, S. and Moessner, R. (1998). Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement*, 130:193–205.

[Bond et al., 1991] Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N. (1991). Excursion set mass functions for hierarchical Gaussian fluctuations. *ApJ*, 379:440–460.

[Bond and Efstathiou, 1987] Bond, J. R. and Efstathiou, G. (1987). The statistics of cosmic background radiation fluctuations. *MNRAS*, 226:655–687.

[Bond et al., 1996] Bond, J. R., Kofman, L., and Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web. *Nature*, 380:603–+.

[Bond and Myers, 1996a] Bond, J. R. and Myers, S. T. (1996a). The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms. *ApJ Sup.*, 103:1.

[Bond and Myers, 1996b] Bond, J. R. and Myers, S. T. (1996b). The Peak-Patch Picture of Cosmic Catalogs. II. Validation. *ApJ Sup.*, 103:41.

[Bond and Myers, 1996c] Bond, J. R. and Myers, S. T. (1996c). The Peak-Patch Picture of Cosmic Catalogs. III. Application to Clusters. *ApJ Sup.*, 103:63.

[Bond et al., 2010] Bond, N. A., Strauss, M. A., and Cen, R. (2010). Crawling the cosmic network: identifying and quantifying filamentary structure. *MNRAS*, 409:156–168.

[Booth and Schaye, 2009] Booth, C. M. and Schaye, J. (2009). Cosmological simulations of the growth of supermassive black holes and feedback from active galactic nuclei: method and tests. *MNRAS*, 398:53–74.

[Bridle and King, 2007] Bridle, S. and King, L. (2007). Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements. *New Journal of Physics*, 9:444.

[Brown et al., 2002] Brown, M. L., Taylor, A. N., Hambly, N. C., and Dye, S. (2002). Measurement of intrinsic alignments in galaxy ellipticities. *MNRAS*, 333:501–509.

[Brunino et al., 2007] Brunino, R., Trujillo, I., Pearce, F. R., and Thomas, P. A. (2007). The orientation of galaxy dark matter haloes around cosmic voids. *MNRAS*, 375:184–190.

[Bryan et al., 1995] Bryan, G. L., Norman, M. L., Stone, J. M., Cen, R., and Ostriker, J. P. (1995). A piecewise parabolic method for cosmological hydrodynamics. *Computer Physics Communications*, 89:149–168.

[Carrasco et al., 2012] Carrasco, J. J. M., Hertzberg, M. P., and Senatore, L. (2012). The Effective Field Theory of Cosmological Large Scale Structures. *ArXiv e-prints*.

[Catelan et al., 2001] Catelan, P., Kamionkowski, M., and Blandford, R. D. (2001). Intrinsic and extrinsic galaxy alignment. *MNRAS*, 320:L7–L13.

[Catelan and Theuns, 1996] Catelan, P. and Theuns, T. (1996). Evolution of the angular momentum of protogalaxies from tidal torques: Zel'dovich approximation. *MNRAS*, 282:436–454.

[Cautun et al., 2013] Cautun, M., van de Weygaert, R., and Jones, B. J. T. (2013). NEXUS: tracing the cosmic web connection. *MNRAS*, 429:1286–1308.

[Centrella and Melott, 1983] Centrella, J. and Melott, A. L. (1983). Three-dimensional simulation of large-scale structure in the universe. *Nature*, 305:196–198.

[Chambers, 1967] Chambers, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika*, 54(3-4):367.

[Chisari et al., 2015] Chisari, N. E., Codis, S., Laigle, C., Dubois, Y., Pichon, C., Devriendt, J., Slyz, A., Miller, L., Gavazzi, R., and Benabed, K. (2015). Intrinsic alignments of galaxies in the Horizon-AGN cosmological hydrodynamical simulation. *ArXiv e-prints*.

[Clerkin et al., 2015] Clerkin, L., Kirk, D., Lahav, O., Abdalla, F. B., and Gaztañaga, E. (2015). A prescription for galaxy biasing evolution as a nuisance parameter. *MNRAS*, 448:1389–1401.

[Codis et al., 2015a] Codis, S., Gavazzi, R., Dubois, Y., Pichon, C., Benabed, K., Desjacques, V., Pogosyan, D., Devriendt, J., and Slyz, A. (2015a). Intrinsic alignment of simulated galaxies in the cosmic web: implications for weak lensing surveys. *MNRAS*, 448:3391–3404.

[Codis et al., 2012] Codis, S., Pichon, C., Devriendt, J., Slyz, A., Pogosyan, D., Dubois, Y., and Sousbie, T. (2012). Connecting the cosmic web to the spin of dark haloes: implications for galaxy formation. *MNRAS*, 427:3320–3336.

[Codis et al., 2015b] Codis, S., Pichon, C., and Pogosyan, D. (2015b). Spin alignments within the cosmic web: a theory of constrained tidal torques near filaments. *MNRAS*, 452:3369–3393.

[Codis et al., 2013] Codis, S., Pichon, C., Pogosyan, D., Bernardeau, F., and Matsubara, T. (2013). Non-Gaussian Minkowski functionals and extrema counts in redshift space. *MNRAS*, 435:531–564.

[Coles, 1988] Coles, P. (1988). Statistical geometry and the microwave background. *MNRAS*, 234:509–531.

[Colombi et al., 2015] Colombi, S., Sousbie, T., Peirani, S., Plum, G., and Suto, Y. (2015). Vlasov versus N-body: the Hénon sphere. *MNRAS*, 450:3724–3741.

[Couchman, 1991] Couchman, H. M. P. (1991). Mesh-refined P3M - A fast adaptive N-body algorithm. *ApJ Let.*, 368:L23–L26.

[Cramér, 1946] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[Crittenden et al., 2001] Crittenden, R. G., Natarajan, P., Pen, U.-L., and Theuns, T. (2001). Spin-induced Galaxy Alignments and Their Implications for Weak-Lensing Measurements. *ApJ*, 559:552–571.

[Crocce et al., 2013] Crocce, M., Castander, F. J., Gaztanaga, E., Fosalba, P., and Carretero, J. (2013). The MICE Grand Challenge Lightcone Simulation II: Halo and Galaxy catalogues. *ArXiv e-prints*.

[Crocce and Scoccimarro, 2006] Crocce, M. and Scoccimarro, R. (2006). Memory of initial conditions in gravitational clustering. *Phys. Rev. D*, 73(6):063520–+.

[Croft and Metzler, 2000] Croft, R. A. C. and Metzler, C. A. (2000). Weak-Lensing Surveys and the Intrinsic Correlation of Galaxy Ellipticities. *ApJ*, 545:561–571.

[Croton et al., 2006] Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., and Yoshida, N. (2006). The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365:11–28.

[Davis et al., 1985] Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394.

[de Lapparent et al., 1986] de Lapparent, V., Geller, M. J., and Huchra, J. P. (1986). A slice of the universe. *ApJ Let.*, 302:L1–L5.

[Dekel and Lahav, 1999] Dekel, A. and Lahav, O. (1999). Stochastic Nonlinear Galaxy Biasing. *ApJ*, 520:24–34.

[Dekel and Rees, 1987] Dekel, A. and Rees, M. J. (1987). Physical mechanisms for biased galaxy formation. *Nature*, 326:455–462.

[Doroshkevich, 1970] Doroshkevich, A. G. (1970). Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, 6:320–330.

[Doroshkevich et al., 1980] Doroshkevich, A. G., Kotok, E. V., Poliudov, A. N., Shandarin, S. F., Sigov, I. S., and Novikov, I. D. (1980). Two-dimensional simulation of the gravitational system dynamics and formation of the large-scale structure of the universe. *MNRAS*, 192:321–337.

[Dressler, 1980] Dressler, A. (1980). Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *ApJ*, 236:351–365.

[Dubois et al., 2012] Dubois, Y., Devriendt, J., Slyz, A., and Teyssier, R. (2012). Self-regulated growth of supermassive black holes by a dual jet-heating active galactic nucleus feedback mechanism: methods, tests and implications for cosmological simulations. *MNRAS*, 420:2662–2683.

[Dubois et al., 2014] Dubois, Y., Pichon, C., Welker, C., Le Borgne, D., Devriendt, J., Laigle, C., Codis, S., Pogosyan, D., Arnouts, S., Benabed, K., Bertin, E., Blaizot, J., Bouchet, F., Cardoso, J.-F., Colombi, S., and de Lapparent, V. (2014). Dancing in the dark: galactic properties trace spin swings along the cosmic web. *MNRAS*, 444:1453–1468.

[Dubois and Teyssier, 2008] Dubois, Y. and Teyssier, R. (2008). On the onset of galactic winds in quiescent star forming galaxies. *A&A*, 477:79–94.

[Efstathiou and Eastwood, 1981a] Efstathiou, G. and Eastwood, J. W. (1981a). On the clustering of particles in an expanding universe. *MNRAS*, 194:503–525.

[Efstathiou and Eastwood, 1981b] Efstathiou, G. and Eastwood, J. W. (1981b). On the clustering of particles in an expanding universe. *MNRAS*, 194:503–525.

[Efstathiou and Jones, 1979] Efstathiou, G. and Jones, B. J. T. (1979). The rotation of galaxies - Numerical investigations of the tidal torque theory. *MNRAS*, 186:133–144.

[Evrard, 1988] Evrard, A. E. (1988). Beyond N-body - 3D cosmological gas dynamics. *MNRAS*, 235:911–934.

[Forero-Romero et al., 2009] Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., and Yepes, G. (2009). A dynamical classification of the cosmic web. *MNRAS*, 396:1815–1824.

[Frenk et al., 1983] Frenk, C. S., White, S. D. M., and Davis, M. (1983). Nonlinear evolution of large-scale structure in the universe. *ApJ*, 271:417–430.

[Fry, 1984] Fry, J. N. (1984). The Galaxy correlation hierarchy in perturbation theory. *ApJ*, 279:499–510.

[Fry, 1996] Fry, J. N. (1996). The Evolution of Bias. *ApJ Let.*, 461:L65.

[Fryxell et al., 2000] Fryxell, B., Olson, K., Ricker, P., Timmes, F. X., Zingale, M., Lamb, D. Q., MacNeice, P., Rosner, R., Truran, J. W., and Tufo, H. (2000). Flash: An adaptive mesh hydrodynamics code for modeling astrophysical thermonuclear flashes. *Astrophysical Journal, Supplement*, 131:273–334.

[Gay et al., 2010] Gay, C., Pichon, C., Le Borgne, D., Teyssier, R., Sousbie, T., and Devriendt, J. (2010). On the filamentary environment of galaxies. *MNRAS*, 404:1801–1816.

[Gay et al., 2012] Gay, C., Pichon, C., and Pogosyan, D. (2012). Non-Gaussian statistics of critical sets in 2D and 3D: Peaks, voids, saddles, genus, and skeleton. *Phys. Rev. D*, 85(2):023011.

[Gaztañaga et al., 2009] Gaztañaga, E., Cabré, A., Castander, F., Crocce, M., and Fosalba, P. (2009). Clustering of luminous red galaxies - III. Baryon acoustic peak in the three-point correlation. *MNRAS*, 399:801–811.

[Geller and Huchra, 1989] Geller, M. J. and Huchra, J. P. (1989). Mapping the universe. *Science*, 246:897–903.

[Gil-Marín et al., 2014] Gil-Marín, H., Wagner, C., Noreña, J., Verde, L., and Percival, W. (2014). Dark matter and halo bispectrum in redshift space: theory and applications. *JCAP* , 12:29.

[González and Padilla, 2010] González, R. E. and Padilla, N. D. (2010). Automated detection of filaments in the large-scale structure of the Universe. *MNRAS*, 407:1449–1463.

[Gott et al., 1989] Gott, III, J. R., Miller, J., Thuan, T. X., Schneider, S. E., Weinberg, D. H., Gammie, C., Polk, K., Vogeley, M., Jeffrey, S., Bhavsar, S. P., Melott, A. L., Giovanelli, R., Hayes, M. P., Tully, R. B., and Hamilton, A. J. S. (1989). The topology of large-scale structure. III - Analysis of observations. *ApJ*, 340:625–646.

[Gott et al., 1990] Gott, III, J. R., Park, C., Juszkiewicz, R., Bies, W. E., Bennett, D. P., Bouchet, F. R., and Stebbins, A. (1990). Topology of microwave background fluctuations - Theory. *ApJ*, 352:1–14.

[Gott et al., 1987] Gott, III, J. R., Weinberg, D. H., and Melott, A. L. (1987). A quantitative approach to the topology of large-scale structure. *ApJ*, 319:1–8.

[Gott et al., 2007] Gott, J. R. I., Colley, W. N., Park, C.-G., Park, C., and Mugnolo, C. (2007). Genus topology of the cosmic microwave background from the WMAP 3-year data. *MNRAS*, 377(4):1668–1678.

[Greggio and Renzini, 1983] Greggio, L. and Renzini, A. (1983). The binary model for type I supernovae - Theoretical rates. *A&A*, 118:217–222.

[Gregory and Thompson, 1978] Gregory, S. A. and Thompson, L. A. (1978). The Coma/A1367 supercluster and its environs. *ApJ*, 222:784–799.

[Gunn and Gott, 1972] Gunn, J. E. and Gott, III, J. R. (1972). On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *ApJ*, 176:1.

[Guo and Jing, 2009] Guo, H. and Jing, Y. P. (2009). Determine the Galaxy Bias Factors on Large Scales Using the Bispectrum Method. *ApJ*, 702:425–432.

[Gurbatov et al., 1989] Gurbatov, S. N., Saichev, A. I., and Shandarin, S. F. (1989). The large-scale structure of the universe in the frame of the model equation of non-linear diffusion. *MNRAS*, 236:385–402.

[Guth, 1981] Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356.

[Guzzo et al., 2008] Guzzo, L., Pierleoni, M., Meneux, B., Branchini, E., Le Fèvre, O., Marinoni, C., Garilli, B., Blaizot, J., De Lucia, G., Pollo, A., and et al. (2008). A test of the nature of cosmic acceleration using galaxy redshift distortions. *Nature*, 451:541–544.

[Guzzo et al., 1997] Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., and Haynes, M. P. (1997). Redshift-Space Distortions and the Real-Space Clustering of Different Galaxy Types. *ApJ*, 489:37–48.

[Haardt and Madau, 1996] Haardt, F. and Madau, P. (1996). Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background. *ApJ*, 461:20–+.

[Hadwiger, 1957] Hadwiger, H. (1957). *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie*. Springer.

[Hahn and Angulo, 2015] Hahn, O. and Angulo, R. E. (2015). An adaptively refined phase-space element method for cosmological simulations and collisionless dynamics. *ArXiv e-prints*.

[Hahn et al., 2007a] Hahn, O., Carollo, C. M., Porciani, C., and Dekel, A. (2007a). The evolution of dark matter halo properties in clusters, filaments, sheets and voids. *MNRAS*, 381:41–51.

[Hahn et al., 2007b] Hahn, O., Porciani, C., Carollo, C. M., and Dekel, A. (2007b). Properties of dark matter haloes in clusters, filaments, sheets and voids. *MNRAS*, 375:489–499.

[Hahn et al., 2009] Hahn, O., Porciani, C., Dekel, A., and Carollo, C. M. (2009). Tidal effects and the environment dependence of halo assembly. *MNRAS*, 398:1742–1756.

[Hahn et al., 2010] Hahn, O., Teyssier, R., and Carollo, C. M. (2010). The large-scale orientations of disc galaxies. *MNRAS*, 405:274–290.

[Hamilton et al., 1986] Hamilton, A. J. S., Gott, III, J. R., and Weinberg, D. (1986). The topology of the large-scale structure of the universe. *ApJ*, 309:1–12.

[Heavens et al., 2000] Heavens, A., Refregier, A., and Heymans, C. (2000). Intrinsic correlation of galaxy shapes: implications for weak lensing measurements. *MNRAS*, 319:649–656.

[Hermit et al., 1996] Hermit, S., Santiago, B. X., Lahav, O., Strauss, M. A., Davis, M., Dressler, A., and Huchra, J. P. (1996). The two-point correlation function and morphological segregation in the Optical Redshift Survey. *MNRAS*, 283:709–720.

[Heymans et al., 2004] Heymans, C., Brown, M., Heavens, A., Meisenheimer, K., Taylor, A., and Wolf, C. (2004). Weak lensing with COMBO-17: estimation and removal of intrinsic alignments. *MNRAS*, 347:895–908.

[Heymans et al., 2006] Heymans, C., White, M., Heavens, A., Vale, C., and van Waerbeke, L. (2006). Potential sources of contamination to weak lensing measurements: constraints from N-body simulations. *MNRAS*, 371:750–760.

[Hidding et al., 2012] Hidding, J., van de Weygaert, R., Vegter, G., and Jones, B. J. T. (2012). Adhesion and the Geometry of the Cosmic Web. *ArXiv e-prints*.

[Hikage et al., 2003] Hikage, C., Schmalzing, J., Buchert, T., Suto, Y., Kayo, I., Taruya, A., Vogeley, M. S., Hoyle, F., Gott, III, J. R., and Brinkmann, J. (2003). Minkowski Functionals of SDSS Galaxies I : Analysis of Excursion Sets. *PASJ*, 55:911–931.

[Hikage et al., 2002] Hikage, C., Suto, Y., Kayo, I., Taruya, A., Matsubara, T., Vogeley, M. S., Hoyle, F., Gott, III, J. R., and Brinkmann, J. (2002). Three-Dimensional Genus Statistics of Galaxies in the SDSS Early Data Release. *PASJ*, 54:707–717.

[Hiramatsu and Taruya, 2009] Hiramatsu, T. and Taruya, A. (2009). Chasing the nonlinear evolution of matter power spectrum with a numerical resummation method: Solution of closure equations. *Phys. Rev. D*, 79(10):103526.

[Hirata et al., 2007] Hirata, C. M., Mandelbaum, R., Ishak, M., Seljak, U., Nichol, R., Pimbblet, K. A., Ross, N. P., and Wake, D. (2007). Intrinsic galaxy alignments from the 2SLAQ and SDSS surveys: luminosity and redshift scalings and implications for weak lensing surveys. *MNRAS*, 381:1197–1218.

[Hirata et al., 2004] Hirata, C. M., Mandelbaum, R., Seljak, U., and et al. (2004). Galaxy-galaxy weak lensing in the Sloan Digital Sky Survey: intrinsic alignments and shear calibration errors. *MNRAS*, 353:529–549.

[Hirata and Seljak, 2004] Hirata, C. M. and Seljak, U. (2004). Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *Phys. Rev. D*, 70(6):063526.

[Hivon et al., 1995] Hivon, E., Bouchet, F. R., Colombi, S., and Juszkiewicz, R. (1995). Redshift distortions of clustering: a Lagrangian approach. *A&A*, 298:643.

[Hoffman et al., 2012] Hoffman, Y., Metuki, O., Yepes, G., Gottlöber, S., Forero-Romero, J. E., Libeskind, N. I., and Knebe, A. (2012). A kinematic classification of the cosmic web. *MNRAS*, 425:2049–2057.

[Hogg et al., 2003] Hogg, D. W., Blanton, M. R., Eisenstein, D. J., Gunn, J. E., Schlegel, D. J., Zehavi, I., Bahcall, N. A., Brinkmann, J., Csabai, I., Schneider, D. P., Weinberg, D. H., and York, D. G. (2003). The Overdensities of Galaxy Environments as a Function of Luminosity and Color. *ApJ Let.*, 585:L5–L9.

[Hoyle, 1949] Hoyle, F. (1949). *Problems of Cosmical Aerodynamics, Central Air Documents, Office, Dayton, OH.* Central Air Documents Office, Dayton, OH.

[Huchra and Geller, 1982] Huchra, J. P. and Geller, M. J. (1982). Groups of galaxies. I - Nearby groups. *ApJ*, 257:423–437.

[Huchra et al., 2012] Huchra, J. P., Macri, L. M., Masters, K. L., Jarrett, T. H., Berlind, P., Calkins, M., Crook, A. C., Cutri, R., Erdoğdu, P., Falco, E., and et al. (2012). The 2MASS Redshift Survey : Description and Data Release. *ApJ Sup.*, 199:26.

[Icke, 1973] Icke, V. (1973). Formation of Galaxies inside Clusters. *A&A*, 27:1.

[Jõeveer et al., 1978] Jõeveer, M., Einasto, J., and Tago, E. (1978). Spatial distribution of galaxies and of clusters of galaxies in the southern galactic hemisphere. *MNRAS*, 185:357–370.

[Jackson, 1972] Jackson, J. C. (1972). A critique of Rees's theory of primordial gravitational radiation. *MNRAS*, 156:1P.

[Jenkins, 2010] Jenkins, A. (2010). Second-order Lagrangian perturbation theory initial conditions for resimulations. *MNRAS*, 403:1859–1872.

[Joachimi and Bridle, 2010] Joachimi, B. and Bridle, S. L. (2010). Simultaneous measurement of cosmology and intrinsic alignments using joint cosmic shear and galaxy number density correlations. *A&A*, 523:A1.

[Joachimi et al., 2015] Joachimi, B., Cacciato, M., Kitching, T. D., Leonard, A., Mandelbaum, R., Schäfer, B. M., Sifón, C., Hoekstra, H., Kiessling, A., Kirk, D., and Rassat, A. (2015). Galaxy alignments: An overview. *ArXiv e-prints*.

[Joachimi et al., 2011] Joachimi, B., Mandelbaum, R., Abdalla, F. B., and Bridle, S. L. (2011). Constraints on intrinsic alignment contamination of weak lensing surveys using the MegaZ-LRG sample. *A&A*, 527:A26.

[Joachimi and Schneider, 2008] Joachimi, B. and Schneider, P. (2008). The removal of shear-ellipticity correlations from the cosmic shear signal via nulling techniques. *A&A*, 488:829–843.

[Joachimi and Schneider, 2010] Joachimi, B. and Schneider, P. (2010). Intrinsic alignment boosting. Direct measurement of intrinsic alignments in cosmic shear data. *A&A*, 517:A4.

[Joachimi et al., 2013a] Joachimi, B., Semboloni, E., Bett, P. E., Hartlap, J., Hilbert, S., Hoekstra, H., Schneider, P., and Schrabback, T. (2013a). Intrinsic galaxy shapes and alignments - I. Measuring and modelling COSMOS intrinsic galaxy ellipticities. *MNRAS*, 431:477–492.

[Joachimi et al., 2013b] Joachimi, B., Semboloni, E., Hilbert, S., Bett, P. E., Hartlap, J., Hoekstra, H., and Schneider, P. (2013b). Intrinsic galaxy shapes and alignments - II. Modelling the intrinsic alignment contamination of weak lensing surveys. *MNRAS*, 436:819–838.

[Jones et al., 2010] Jones, B. J. T., van de Weygaert, R., and Aragón-Calvo, M. A. (2010). Fossil evidence for spin alignment of Sloan Digital Sky Survey galaxies in filaments. *MNRAS*, 408:897–918.

[Jost, 2008] Jost, J. (2008). *Riemannian Geometry and Geometric Analysis, Fifth Edition*. Berlin ; New York : Springer, c2008.

[Juszkiewicz et al., 1995] Juszkiewicz, R., Weinberg, D. H., Amsterdamski, P., Chodorowski, M., and Bouchet, F. (1995). Weakly nonlinear Gaussian fluctuations and the edgeworth expansion. *ApJ*, 442:39–56.

[Kac, 1943] Kac, M. (1943). On the average number of real roots of a random algebraic equation. *Bull. Am. Math. Soc.*, 49:938–+.

[Kaiser, 1984] Kaiser, N. (1984). On the spatial correlations of Abell clusters. *ApJ Let.*, 284:L9–L12.

[Kaiser, 1987] Kaiser, N. (1987). Clustering in real space and in redshift space. *MNRAS*, 227:1–21.

[Kauffmann et al., 1999] Kauffmann, G., Colberg, J. M., Diaferio, A., and White, S. D. M. (1999). Clustering of galaxies in a hierarchical universe - I. Methods and results at z=0. *MNRAS*, 303:188–206.

[Kendall and Stuart, 1958] Kendall, M. G. and Stuart, A. (1958). *The Advanced Theory of Statistics*, volume 1. London: Griffin.

[Kennicutt, 1998] Kennicutt, Jr., R. C. (1998). The Global Schmidt Law in Star-forming Galaxies. *ApJ*, 498:541–+.

[Khandai et al., 2015] Khandai, N., Di Matteo, T., Croft, R., Wilkins, S., Feng, Y., Tucker, E., DeGraf, C., and Liu, M.-S. (2015). The MassiveBlack-II simulation: the evolution of haloes and galaxies to z = 0. *MNRAS*, 450:1349–1374.

[Kiessling et al., 2015] Kiessling, A., Cacciato, M., Joachimi, B., Kirk, D., Kitching, T. D., Leonard, A., Mandelbaum, R., Schäfer, B. M., Sifón, C., Brown, M. L., and Rassat, A. (2015). Galaxy alignments: Theory, modelling and simulations. *ArXiv e-prints*.

[Kilbinger et al., 2013] Kilbinger, M., Fu, L., Heymans, C., Simpson, F., Benjamin, J., Erben, T., Harnois-Déraps, J., Hoekstra, H., Hildebrandt, H., Kitching, T. D., and et al. (2013). CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing. *MNRAS*, 430:2200–2220.

[Kim et al., 2011] Kim, J., Park, C., Rossi, G., Lee, S. M., and Gott, III, J. R. (2011). The New Horizon Run Cosmological N-Body Simulations. *Journal of Korean Astronomical Society*, 44:217–234.

[Kirk et al., 2010] Kirk, D., Bridle, S., and Schneider, M. (2010). The impact of intrinsic alignments: cosmological constraints from a joint analysis of cosmic shear and galaxy survey data. *MNRAS*, 408:1502–1515.

[Klypin and Shandarin, 1993] Klypin, A. and Shandarin, S. F. (1993). Percolation technique for galaxy clustering. *ApJ*, 413:48–58.

[Klypin et al., 2011] Klypin, A. A., Trujillo-Gomez, S., and Primack, J. (2011). Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. *ApJ*, 740:102.

[Kofman et al., 1990] Kofman, L., Pogosian, D., and Shandarin, S. (1990). Structure of the universe in the two-dimensional model of adhesion. *MNRAS*, 242:200–208.

[Komatsu et al., 2011] Komatsu, E., Smith, K. M., Dunkley, J., and et al. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *ApJ Sup.*, 192:18–+.

[Kovač et al., 2014] Kovač, K., Lilly, S. J., Knobel, C., Bschorr, T. J., Peng, Y., Carollo, C. M., Contini, T., Kneib, J.-P., Le Févre, O., Mainieri, V., and et al. (2014). zCOSMOS 20k: satellite galaxies are the main drivers of environmental effects in the galaxy population at least to z = 0.7. *MNRAS*, 438:717–738.

[Krumholz and Tan, 2007] Krumholz, M. R. and Tan, J. C. (2007). Slow Star Formation in Dense Gas: Evidence and Implications. *ApJ*, 654:304–315.

[Laigle et al., 2015] Laigle, C., Pichon, C., Codis, S., Dubois, Y., Le Borgne, D., Pogosyan, D., Devriendt, J., Peirani, S., Prunet, S., Rouberol, S., Slyz, A., and Sousbie, T. (2015). Swirling around filaments: are large-scale structure vortices spinning up dark haloes? *MNRAS*, 446:2744–2759.

[Laureijs et al., 2011] Laureijs, R., Amiaux, J., Arduini, S., Augueres, J. ., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., and et al. (2011). Euclid Definition Study Report. *ArXiv e-prints*.

[Lee and Erdogdu, 2007] Lee, J. and Erdogdu, P. (2007). The Alignments of the Galaxy Spins with the Real-Space Tidal Field Reconstructed from the 2MASS Redshift Survey. *ApJ*, 671:1248–1255.

[Lee and Lee, 2008] Lee, J. and Lee, B. (2008). The Variation of Galaxy Morphological Type with Environmental Shear. *ApJ*, 688:78–84.

[Lee and Li, 2008] Lee, J. and Li, C. (2008). Connecting the Physical Properties of Galaxies with the Overdensity and Tidal Shear of the Large-Scale Environment. *ArXiv e-prints*.

[Lee and Pen, 2000] Lee, J. and Pen, U.-L. (2000). Cosmic Shear from Galaxy Spins. *ApJ Let.*, 532:L5–L8.

[Lee and Pen, 2001] Lee, J. and Pen, U.-L. (2001). Galaxy Spin Statistics and Spin-Density Correlation. *ApJ*, 555:106–124.

[Lee and Pen, 2002] Lee, J. and Pen, U.-L. (2002). Detection of Galaxy Spin Alignments in the Point Source Catalog Redshift Survey Shear Field. *ApJ Let.*, 567:L111–L114.

[Lee and Pen, 2008] Lee, J. and Pen, U.-L. (2008). The Nonlinear Evolution of Galaxy Intrinsic Alignments. *ApJ*, 681:798–805.

[Leitherer et al., 2010] Leitherer, C., Ortiz Otálvaro, P. A., Bresolin, F., Kudritzki, R.-P., Lo Faro, B., Pauldrach, A. W. A., Pettini, M., and Rix, S. A. (2010). A Library of Theoretical Ultraviolet Spectra of Massive, Hot Stars for Evolutionary Synthesis. *ApJ Sup.*, 189:309–335.

[Leitherer et al., 1999] Leitherer, C., Schaerer, D., Goldader, J. D., and et al. (1999). Starburst99: Synthesis Models for Galaxies with Active Star Formation. *ApJ Sup.*, 123:3–40.

[Lemson, 1993] Lemson, G. (1993). Dynamical Effects of the Cosmological Constant - the Evolution of Aspherical Structures. *MNRAS*, 263:913.

[Lesgourgues et al., 2009] Lesgourgues, J., Matarrese, S., Pietroni, M., and Riotto, A. (2009). Non-linear power spectrum including massive neutrinos: the time-RG flow approach. *JCAP* , 6:17.

[Libeskind et al., 2013] Libeskind, N. I., Hoffman, Y., Steinmetz, M., Gottlöber, S., Knebe, A., and Hess, S. (2013). Cosmic vorticity and the origin halo spins. *ApJ Let.*, 766:L15.

[Lin et al., 1965] Lin, C. C., Mestel, L., and Shu, F. H. (1965). The Gravitational Collapse of a Uniform Spheroid. *ApJ*, 142:1431.

[Lokas et al., 1995] Lokas, E. L., Juszkiewicz, R., Weinberg, D. H., and Bouchet, F. R. (1995). Kurtosis of large-scale cosmic fields. *MNRAS*, 274:730–744.

[Lumsden et al., 1989] Lumsden, S. L., Heavens, A. F., and Peacock, J. A. (1989). The clustering of peaks in a random Gaussian field. *MNRAS*, 238:293–318.

[Lynden-Bell, 1964] Lynden-Bell, D. (1964). On Large-Scale Instabilities during Gravitational Collapse and the Evolution of Shrinking Maclaurin Spheroids. *ApJ*, 139:1195.

[Ma and Fry, 2000] Ma, C.-P. and Fry, J. N. (2000). Deriving the Nonlinear Cosmological Power Spectrum and Bispectrum from Analytic Dark Matter Halo Profiles and Mass Functions. *ApJ*, 543:503–513.

[Mandelbaum et al., 2011] Mandelbaum, R., Blake, C., Bridle, and et al. (2011). The WiggleZ Dark Energy Survey: direct constraints on blue galaxy intrinsic alignments at intermediate redshifts. *MNRAS*, 410:844–859.

[Mandelbaum et al., 2006] Mandelbaum, R., Hirata, C. M., Ishak, M., Seljak, U., and Brinkmann, J. (2006). Detection of large-scale intrinsic ellipticity-density correlation from the Sloan Digital Sky Survey and implications for weak lensing surveys. *MNRAS*, 367:611–626.

[Mann et al., 1998] Mann, R. G., Peacock, J. A., and Heavens, A. F. (1998). Eulerian bias and the galaxy density field. *MNRAS*, 293:209.

[Matsubara, 1994] Matsubara, T. (1994). Analytic expression of the genus in a weakly non-Gaussian field induced by gravity. *ApJ Let.*, 434:L43–L46.

[Matsubara, 1996] Matsubara, T. (1996). Statistics of Isodensity Contours in Redshift Space. *ApJ*, 457:13.

[Matsubara, 1999] Matsubara, T. (1999). Stochasticity of Bias and Nonlocality of Galaxy Formation: Linear Scales. *ApJ*, 525:543–553.

[Matsubara, 2010] Matsubara, T. (2010). Analytic Minkowski functionals of the cosmic microwave background: Second-order non-Gaussianity with bispectrum and trispectrum. *Phys. Rev. D*, 81(8):083505–+.

[Mecke et al., 1994] Mecke, K. R., Buchert, T., and Wagner, H. (1994). Robust morphological measures for large-scale structure in the Universe. *A&A*, 288:697–704.

[Melott et al., 1989] Melott, A. L., Cohen, A. P., Hamilton, A. J. S., Gott, III, J. R., and Weinberg, D. H. (1989). Topology of large-scale structure. IV - Topology in two dimensions. *ApJ*, 345:618–626.

[Melott et al., 1988] Melott, A. L., Weinberg, D. H., and Gott, III, J. R. (1988). The topology of large-scale structure. II - Nonlinear evolution of Gaussian models. *ApJ*, 328:50–68.

[Metuki et al., 2015] Metuki, O., Libeskind, N. I., Hoffman, Y., Crain, R. A., and Theuns, T. (2015). Galaxy properties and the cosmic web in simulations. *MNRAS*, 446:1458–1468.

[Mo and White, 1996] Mo, H. J. and White, S. D. M. (1996). An analytic model for the spatial clustering of dark matter haloes. *MNRAS*, 282:347–361.

[Musso and Sheth, 2012] Musso, M. and Sheth, R. K. (2012). One step beyond: the excursion set approach with correlated steps. *MNRAS*, 423:L102–L106.

[Navarro et al., 2004] Navarro, J. F., Abadi, M. G., and Steinmetz, M. (2004). Tidal Torques and the Orientation of Nearby Disk Galaxies. *ApJ Let.*, 613:L41–L44.

[Neyrinck, 2012] Neyrinck, M. C. (2012). Origami constraints on the initial-conditions arrangement of dark-matter caustics and streams. *MNRAS*, 427:494–501.

[Nishimichi et al., 2010] Nishimichi, T., Taruya, A., Koyama, K., and Sabiu, C. (2010). Scale dependence of halo bispectrum from non-Gaussian initial conditions in cosmological N-body simulations. *JCAP* , 7:2.

[Norberg et al., 2002] Norberg, P., Baugh, C. M., Hawkins, E., Maddox, S., Madgwick, D., Lahav, O., Cole, S., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., and et al. (2002). The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type. *MNRAS*, 332:827–838.

[Novikov et al., 2006] Novikov, D., Colombi, S., and Doré, O. (2006). Skeleton as a probe of the cosmic web: the two-dimensional case. *MNRAS*, 366:1201–1216.

[Nusser and Davis, 1994] Nusser, A. and Davis, M. (1994). On the prediction of velocity fields from redshift space galaxy samples. *ApJ Let.*, 421:L1–L4.

[Ocvirk et al., 2008] Ocvirk, P., Pichon, C., and Teyssier, R. (2008). Bimodal gas accretion in the Horizon-MareNostrum galaxy formation simulation. *MNRAS*, 390:1326–1338.

[Oemler, 1974] Oemler, Jr., A. (1974). The Systematic Properties of Clusters of Galaxies. Photometry of 15 Clusters. *ApJ*, 194:1–20.

[O'Shea et al., 2004] O'Shea, B. W., Bryan, G., Bordner, J., Norman, M. L., Abel, T., Harkness, R., and Kritsuk, A. (2004). Introducing Enzo, an AMR Cosmology Application. *ArXiv Astrophysics e-prints*.

[Padmanabhan, 1993] Padmanabhan, T. (1993). *Structure Formation in the Universe*.

[Paranjape et al., 2012] Paranjape, A., Lam, T. Y., and Sheth, R. K. (2012). Halo abundances and counts-in-cells: the excursion set approach with correlated steps. *MNRAS*, 420:1429–1441.

[Park et al., 2005] Park, C., Choi, Y.-Y., Vogeley, M. S., Gott, J. R. I., Kim, J., Hikage, C., Matsubara, T., Park, M.-G., Suto, Y., and Weinberg, D. H. (2005). Topology Analysis of the Sloan Digital Sky Survey. I. Scale and Luminosity Dependence. *ApJ*, 633:11–22.

[Paz et al., 2008] Paz, D. J., Stasyszyn, F., and Padilla, N. D. (2008). Angular momentum-large-scale structure alignments in ΛCDM models and the SDSS. *MNRAS*, 389:1127–1136.

[Peacock, 1999] Peacock, J. A. (1999). *Cosmological Physics*.

[Peacock et al., 2001] Peacock, J. A., Cole, S., Norberg, P., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R. D., Colless, M., Collins, C., Couch, W., and et al. (2001). A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. *Nature*, 410:169–173.

[Peacock and Heavens, 1990] Peacock, J. A. and Heavens, A. F. (1990). Alternatives to the Press-Schechter cosmological mass function. *MNRAS*, 243:133–143.

[Peacock and Smith, 2000] Peacock, J. A. and Smith, R. E. (2000). Halo occupation numbers and galaxy bias. *MNRAS*, 318:1144–1156.

[Peebles, 1969] Peebles, P. J. E. (1969). Origin of the Angular Momentum of Galaxies. *ApJ*, 155:393–+.

[Peebles, 1970] Peebles, P. J. E. (1970). Structure of the Coma Cluster of Galaxies. *AJ*, 75:13.

[Peebles, 1980] Peebles, P. J. E. (1980). *The large-scale structure of the universe*.

[Pen, 1998] Pen, U.-L. (1998). Reconstructing Nonlinear Stochastic Bias from Velocity Space Distortions. *ApJ*, 504:601–606.

[Percival et al., 2004] Percival, W. J., Burkey, D., Heavens, A., Taylor, A., Cole, S., Peacock, J. A., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R., and et al. (2004). The 2dF Galaxy Redshift Survey: spherical harmonics analysis of fluctuations in the final catalogue. *MNRAS*, 353:1201–1218.

[Percival et al., 2007] Percival, W. J., Cole, S., Eisenstein, D. J., Nichol, R. C., Peacock, J. A., Pope, A. C., and Szalay, A. S. (2007). Measuring the Baryon Acoustic Oscillation scale using the Sloan Digital Sky Survey and 2dF Galaxy Redshift Survey. *MNRAS*, 381:1053–1066.

[Perlmutter et al., 1999] Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., and et al. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517:565–586.

[Pietroni, 2008] Pietroni, M. (2008). Flowing with time: a new approach to non-linear cosmological perturbations. *JCAP* , 10:36.

[Planck Collaboration et al., 2014a] Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014a). Planck 2013 results. XII. Diffuse component separation. *A&A*, 571:A12.

[Planck Collaboration et al., 2014b] Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014b). Planck 2013 results. XV. CMB power spectra and likelihood. *A&A*, 571:A15.

[Planck Collaboration et al., 2014c] Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014c). Planck 2013 results. XVI. Cosmological parameters. *A&A*, 571:A16.

[Planck Collaboration et al., 2014d] Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014d). Planck 2013 results. XXIII. Isotropy and statistics of the CMB. *A&A*, 571:A23.

[Planck Collaboration et al., 2014e] Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014e). Planck 2013 results. XXIV. Constraints on primordial non-Gaussianity. *A&A*, 571:A24.

[Platen et al., 2007] Platen, E., van de Weygaert, R., and Jones, B. J. T. (2007). A cosmic watershed: the WVF void detection technique. *MNRAS*, 380:551–570.

[Platen et al., 2008] Platen, E., van de Weygaert, R., and Jones, B. J. T. (2008). Alignment of voids in the cosmic web. *MNRAS*, 387:128–136.

[Pogosyan et al., 1998] Pogosyan, D., Bond, J. R., Kofman, L., and Wadsley, J. (1998). Cosmic Web: Origin and Observables. In S. Colombi, Y. Mellier, & B. Raban, editor, *Wide Field Surveys in Cosmology*, page 61.

[Pogosyan et al., 2009a] Pogosyan, D., Gay, C., and Pichon, C. (2009a). Invariant joint distribution of a stationary random field and its derivatives: Euler characteristic and critical point counts in 2 and 3D. *Phys. Rev. D*, 80(8):081301–+.

[Pogosyan et al., 2009b] Pogosyan, D., Pichon, C., Gay, C., Prunet, S., Cardoso, J. F., Sousbie, T., and Colombi, S. (2009b). The local theory of the cosmic skeleton. *MNRAS*, 396:635–667.

[Porciani et al., 2002a] Porciani, C., Dekel, A., and Hoffman, Y. (2002a). Testing tidal-torque theory - I. Spin amplitude and direction. *MNRAS*, 332:325–338.

[Porciani et al., 2002b] Porciani, C., Dekel, A., and Hoffman, Y. (2002b). Testing tidal-torque theory - II. Alignment of inertia and shear and the characteristics of protohaloes. *MNRAS*, 332:339–351.

[Press and Schechter, 1974] Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438.

[Rasera and Teyssier, 2006] Rasera, Y. and Teyssier, R. (2006). The history of the baryon budget. Cosmic logistics in a hierarchical universe. *A&A*, 445:1–27.

[Rice, 1944] Rice, S. O. (1944). Mathematical Analysis of Random Noise. *Bell Systems Tech. J., Volume 23, p. 282-332*, 23:282–332.

[Rice, 1945] Rice, S. O. (1945). Mathematical Analysis of Random Noise-Conclusion. *Bell Systems Tech. J., Volume 24, p. 46-156*, 24:41–156.

[Riess et al., 1998] Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., and et al. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038.

[Rubin et al., 1980] Rubin, V. C., Ford, W. K. J., and . Thonnard, N. (1980). Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *ApJ*, 238:471–487.

[Ryden, 1988] Ryden, B. S. (1988). The area of isodensity contours as a measure of large-scale structure. *ApJ Let.*, 333:L41–L44.

[Ryden et al., 1989] Ryden, B. S., Melott, A. L., Craig, D. A., Gott, III, J. R., Weinberg, D. H., Scherrer, R. J., Bhavsar, S. P., and Miller, J. M. (1989). The area of isodensity contours in cosmological models and galaxy surveys. *ApJ*, 340:647–660.

[Salpeter, 1955] Salpeter, E. E. (1955). The Luminosity Function and Stellar Evolution. *ApJ*, 121:161.

[Sargent and Turner, 1977] Sargent, W. L. W. and Turner, E. L. (1977). A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. *ApJ Let.*, 212:L3–L7.

[Scannapieco et al., 2012] Scannapieco, C., Wadepuhl, M., Parry, O. H., Navarro, J. F., Jenkins, A., Springel, V., Teyssier, R., Carlson, E., Couchman, H. M. P., Crain, R. A., and et al. (2012). The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation. *MNRAS*, 423:1726–1749.

[Schaefer, 2009] Schaefer, B. M. (2009). Galactic Angular Momenta and Angular Momentum Correlations in the Cosmological Large-Scale Structure. *International Journal of Modern Physics D*, 18:173–222.

[Schäfer and Merkel, 2012] Schäfer, B. M. and Merkel, P. M. (2012). Galactic angular momenta and angular momentum couplings in the large-scale structure. *MNRAS*, 421:2751–2762.

[Schaye et al., 2015] Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., and et al. (2015). The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *MNRAS*, 446:521–554.

[Scherrer and Bertschinger, 1991] Scherrer, R. J. and Bertschinger, E. (1991). Statistics of primordial density perturbations from discrete seed masses. *ApJ*, 381:349–360.

[Schneider and Bridle, 2010] Schneider, M. D. and Bridle, S. (2010). A halo model for intrinsic alignments of galaxy ellipticities. *MNRAS*, 402:2127–2139.

[Schneider et al., 2012] Schneider, M. D., Frenk, C. S., and Cole, S. (2012). The shapes and alignments of dark matter halos. *JCAP* , 5:30.

[Schneider et al., 2002] Schneider, P., van Waerbeke, L., Kilbinger, M., and Mellier, Y. (2002). Analysis of two-point statistics of cosmic shear. I. Estimators and covariances. *A&A*, 396:1–19.

[Scoccimarro, 2004] Scoccimarro, R. (2004). Redshift-space distortions, pairwise velocities, and nonlinearities. *Phys. Rev. D*, 70(8):083007.

[Scoccimarro et al., 1998] Scoccimarro, R., Colombi, S., Fry, J. N., Frieman, J. A., Hivon, E., and Melott, A. (1998). Nonlinear Evolution of the Bispectrum of Cosmological Perturbations. *ApJ*, 496:586.

[Scoccimarro et al., 1999] Scoccimarro, R., Couchman, H. M. P., and Frieman, J. A. (1999). The Bispectrum as a Signature of Gravitational Instability in Redshift Space. *ApJ*, 517:531–540.

[Scoccimarro et al., 2001] Scoccimarro, R., Sheth, R. K., Hui, L., and Jain, B. (2001). How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *ApJ*, 546:20–34.

[Seljak, 2000] Seljak, U. (2000). Analytic model for galaxy and dark matter clustering. *MNRAS*, 318:203–213.

[Shakura and Sunyaev, 1973] Shakura, N. I. and Sunyaev, R. A. (1973). Black holes in binary systems. Observational appearance. *A&A*, 24:337–355.

[Shandarin, 1991] Shandarin, S. (1991). Theory of Adhesion for Large-Scale Structure. In Latham, D. W. and da Costa, L. A. N., editors, *Large-scale Structures and Peculiar Motions in the Universe*, volume 15 of *Astronomical Society of the Pacific Conference Series*, page 189.

[Sheth and Tormen, 2002] Sheth, R. K. and Tormen, G. (2002). An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. *MNRAS*, 329:61–75.

[Sousbie, 2011] Sousbie, T. (2011). The persistent cosmic web and its filamentary structure - I. Theory and implementation. *MNRAS*, 414:350–383.

[Sousbie et al., 2009] Sousbie, T., Colombi, S., and Pichon, C. (2009). The fully connected N-dimensional skeleton: probing the evolution of the cosmic web. *MNRAS*, 393:457–477.

[Sousbie et al., 2008] Sousbie, T., Pichon, C., Colombi, S., and Pogosyan, D. (2008). The 3D skeleton: tracing the filamentary structure of the Universe. *MNRAS*, 383(4):1655–1670.

[Sousbie et al., 2011] Sousbie, T., Pichon, C., and Kawahara, H. (2011). The persistent cosmic web and its filamentary structure - II. Illustrations. *MNRAS*, 414:384–403.

[Spergel et al., 2003] Spergel, D. N., Verde, L., Peiris, H. V., Komatsu, E., Nolta, M. R., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., and et al. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *ApJ Sup.*, 148:175–194.

[Springel, 2005] Springel, V. (2005). The cosmological simulation code GADGET-2. *MNRAS*, 364:1105–1134.

[Springel, 2010] Springel, V. (2010). E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *MNRAS*, 401:791–851.

[Springel et al., 2005] Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., and et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636.

[Springel et al., 2001] Springel, V., White, S. D. M., Tormen, G., and Kauffmann, G. (2001). Populating a cluster of galaxies - I. Results at [formmu2]z=0. *MNRAS*, 328:726–750.

[Stoica et al., 2005] Stoica, R. S., Martínez, V. J., Mateu, J., and Saar, E. (2005). Detection of cosmic filaments using the Candy model. *A&A*, 434:423–432.

[Sugerman et al., 2000] Sugerman, B., Summers, F. J., and Kamionkowski, M. (2000). Testing linear-theory predictions of galaxy formation. *MNRAS*, 311:762–780.

[Sutherland and Dopita, 1993] Sutherland, R. S. and Dopita, M. A. (1993). Cooling functions for low-density astrophysical plasmas. *ApJ Sup.*, 88:253–327.

[Taruya et al., 2012] Taruya, A., Bernardeau, F., Nishimichi, T., and Codis, S. (2012). Direct and fast calculation of regularized cosmological power spectrum at two-loop order. *Phys. Rev. D*, 86(10):103528.

[Taruya and Hiramatsu, 2008] Taruya, A. and Hiramatsu, T. (2008). A Closure Theory for Nonlinear Evolution of Cosmological Power Spectra. *ApJ*, 674:617–635.

[Taruya et al., 2010] Taruya, A., Nishimichi, T., and Saito, S. (2010). Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory. *Phys. Rev. D*, 82(6):063522.

[Tegmark et al., 2004] Tegmark, M., Blanton, M. R., Strauss, M. A., Hoyle, F., Schlegel, D., Scoccimarro, R., Vogeley, M. S., Weinberg, D. H., Zehavi, I., Berlind, A., and et al. (2004). The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. *ApJ*, 606:702–740.

[Tegmark and Peebles, 1998] Tegmark, M. and Peebles, P. J. E. (1998). The Time Evolution of Bias. *ApJ Let.*, 500:L79–L82.

[Tempel and Libeskind, 2013] Tempel, E. and Libeskind, N. I. (2013). Galaxy Spin Alignment in Filaments and Sheets: Observational Evidence. *ApJ Let.*, 775:L42.

[Tempel et al., 2011] Tempel, E., Saar, E., Liivamägi, L. J., Tamm, A., Einasto, J., Einasto, M., and Müller, V. (2011). Galaxy morphology, luminosity, and environment in the SDSS DR7. *A&A*, 529:A53.

[Tempel et al., 2014] Tempel, E., Stoica, R. S., Martínez, V. J., Liivamägi, L. J., Castellan, G., and Saar, E. (2014). Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS. *MNRAS*, 438:3465–3482.

[Tempel et al., 2013] Tempel, E., Stoica, R. S., and Saar, E. (2013). Evidence for spin alignment of spiral and elliptical/S0 galaxies in filaments. *MNRAS*, 428(2):1827–1836.

[Tenneti et al., 2015] Tenneti, A., Singh, S., Mandelbaum, R., Matteo, T. D., Feng, Y., and Khandai, N. (2015). Intrinsic alignments of galaxies in the MassiveBlack-II simulation: analysis of two-point statistics. *MNRAS*, 448:3522–3544.

[Teyssier, 2002] Teyssier, R. (2002). Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *A&A*, 385:337–364.

[Teyssier et al., 2009] Teyssier, R., Pires, S., Prunet, S., Aubert, D., Pichon, C., Amara, A., Benabed, K., Colombi, S., Refregier, A., and Starck, J.-L. (2009). Full-sky weak-lensing simulation with 70 billion particles. *A&A*, 497:335–341.

[Troxel and Ishak, 2015] Troxel, M. A. and Ishak, M. (2015). The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. *Phys. Rep.*, 558:1–59.

[Trujillo et al., 2006] Trujillo, I., Carretero, C., and Patiri, S. G. (2006). Detection of the Effect of Cosmological Large-Scale Structure on the Orientation of Galaxies. *ApJ Let.*, 640:L111–L114.

[Vogelsberger et al., 2014] Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Bird, S., Nelson, D., and Hernquist, L. (2014). Properties of galaxies reproduced by a hydrodynamic simulation. *Nature*, 509:177–182.

[Wadsley et al., 2004] Wadsley, J. W., Stadel, J., and Quinn, T. (2004). Gasoline: a flexible, parallel implementation of TreeSPH. *New Astronomy*, 9:137–158.

[Watanabe and Inagaki, 1991] Watanabe, T. and Inagaki, S. (1991). Peculiar velocity field around superclusters - The effect of asphericity. *PASJ*, 43:413–426.

[Weinberg et al., 1987] Weinberg, D. H., Gott, III, J. R., and Melott, A. L. (1987). The topology of large-scale structure. I - Topology and the random phase hypothesis. *ApJ*, 321:2–27.

[White, 1996] White, S. (1996). Cosmology and large scale structure. *Proceedings of Les Houches Summer School, R. Schaeffer et al., editors,(Elsevier, Amsterdam)*.

[White, 1976] White, S. D. M. (1976). The dynamics of rich clusters of galaxies. *MNRAS*, 177:717–733.

[White, 1984] White, S. D. M. (1984). Angular momentum growth in protogalaxies. *ApJ*, 286:38–41.

[White and Silk, 1979] White, S. D. M. and Silk, J. (1979). The growth of aspherical structure in the universe - Is the local supercluster an unusual system. *ApJ*, 231:1–9.

[White et al., 1988] White, S. D. M., Tully, R. B., and Davis, M. (1988). Clustering bias in the nearby galaxies catalog and in cold dark matter models. *ApJ Let.*, 333:L45–L49.

[Yan et al., 2012] Yan, H., Fan, Z., and White, S. D. M. (2012). On the Tidal Dependence of Galaxy Properties. *ArXiv e-prints*.

[Zeldovich et al., 1982] Zeldovich, I. B., Einasto, J., and Shandarin, S. F. (1982). Giant voids in the universe. *Nature*, 300:407–413.

[Zel'dovich, 1970] Zel'dovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *A&A*, 5:84–89.

[Zhang et al., 2009] Zhang, Y., Yang, X., Faltenbacher, A., Springel, V., Lin, W., and Wang, H. (2009). The Spin and Orientation of Dark Matter Halos Within Cosmic Filaments. *ApJ*, 706:747–761.

[Zhang et al., 2013] Zhang, Y., Yang, X., Wang, H., Wang, L., Mo, H. J., and van den Bosch, F. C. (2013). Alignments of Galaxies within Cosmic Filaments from SDSS DR7. *ApJ*, 779:160.

[Zwicky, 1933] Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127.

A Other publications

In this appendix are gathered some publications of mine that are referred to in the main text but for which I am not one of the leading authors. This includes a work led by Clotilde Laigle about vorticity generation within the cosmic web (see paper A.1), a paper by Yohan Dubois about the spin of galaxies in the Horizon-AGN *simulation (see paper A.2) and finally an article with Atsushi Taruya, Francis Bernardeau and Takahiro Nishimichi about the density power spectrum at two-loop in RegPT (see paper A.3).*

Publications

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **446,** 2744–2759 (2015) doi:10.1093/mnras/stu2289

Swirling around filaments: are large-scale structure vortices spinning up dark haloes?

C. Laigle,[1⋆] C. Pichon,[1,2] S. Codis,[1] Y. Dubois,[1] D. Le Borgne,[1] D. Pogosyan,[3] J. Devriendt,[4] S. Peirani,[1] S. Prunet,[1] S. Rouberol,[1] A. Slyz[5] and T. Sousbie[1]

[1]*Institut d'Astrophysique de Paris & UPMC (UMR 7095), 98 bis boulevard Arago, F-75014 Paris, France*
[2]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3]*Department of Physics, University of Alberta, 11322-89 Avenue, Edmonton, Alberta T6G 2G7, Canada*
[4]*Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*
[5]*KITP Kohn Hall-4030, University of California, Santa Barbara, CA 93106-4030, USA*

Accepted 2014 October 29. Received 2014 October 22; in original form 2013 October 13

ABSTRACT

The kinematic analysis of dark matter and hydrodynamical simulations suggests that the vorticity in large-scale structure is mostly confined to, and predominantly aligned with, their filaments, with an excess of probability of 20 per cent to have the angle between vorticity and filaments direction lower than 60° relative to random orientations. The cross-sections of these filaments are typically partitioned into four quadrants with opposite vorticity sign, arising from multiple flows, originating from neighbouring walls. The spins of haloes embedded within these filaments are consistently aligned with this vorticity for any halo mass, with a stronger alignment for the most massive structures up to an excess of probability of 165 per cent. The global geometry of the flow within the cosmic web is therefore qualitatively consistent with a spin acquisition for smaller haloes induced by this large-scale coherence, as argued in Codis et al. In effect, secondary anisotropic infall (originating from the vortex-rich filament within which these lower-mass haloes form) dominates the angular momentum budget of these haloes. The transition mass from alignment to orthogonality is related to the size of a given multi-flow region with a given polarity. This transition may be reconciled with the standard tidal torque theory if the latter is augmented so as to account for the larger scale anisotropic environment of walls and filaments.

Key words: methods: numerical – galaxies: formation – galaxies: haloes – large-scale structure of Universe.

1 INTRODUCTION

The standard paradigm of galaxy formation addresses the acquisition of spin via the so-called Tidal Torque Theory (TTT; Hoyle 1949; Peebles 1969; Doroshkevich 1970; White 1984) for which collapsing protogalaxies acquire their spin because of a misalignment between their inertia tensor and their (local) tidal tensor. There is ample evidence that for massive (quasi-linear) clusters, TTT provides a sound theoretical framework in which to describe angular momentum (AM) acquisition during the linear phase of structure formation. Conversely, lighter non-linear structures undergo significant drift within the large-scale tidal field, and move some distance away from their original Lagrangian patch (see e.g. Schaefer 2009, for a review). Over the last 10 years, numerical simulations as well as theoretical consideration (Birnboim & Dekel 2003; Katz et al. 2003; Kereš et al. 2005; Ocvirk, Pichon & Teyssier 2008) have accumulated evidence that the intricate cosmic web plays a critical role in the process of forming high-redshift galaxies. In the initial phase of galaxy formation, the condition of the intergalactic medium leads to essentially isothermal shocks. Hence cold gas follows closely the cosmic web while radiating away the thermal energy gained by the extraction of kinetic energy every time its trajectory dictates the formation of a shock.

The dynamical relevance of the anisotropy of the cosmic web for galaxy formation may have been partially underestimated given the small mass involved (in contrast to the mass in peaks). Indeed, spherical collapse and Press–Schechter theory have been quite successful at explaining the mass function of galaxies (Press & Schechter 1974). On the other hand, the morphology of galaxies, arguably a secondary feature, is controlled at high redshift by their spin (see e.g. Dubois et al. 2012) and is very likely driven by later infall of AM-rich gas. In turn, the critical ingredient must therefore be the anisotropy of such infall, driven by its dynamics within the cosmic web, which differ significantly (via the hitherto mentioned shocks) from that of the dark matter (DM), since cold flows advect the AM

⋆ E-mail: laigle@iap.fr

they acquired as they formed during the early phase of large-scale structure (LSS) formation. A paradigm for the acquisition of disc AM via filamentary flows was recently proposed by Pichon et al. (2011) which found a closer connection between the 3D geometry and dynamics of the neighbouring cosmic web and the properties of embedded dark haloes and galaxies than originally suggested by the standard hierarchical formation paradigm (see also Prieto, Jimenez & Haiman 2013; Stewart et al. 2013). At these scales, in the surrounding asymmetric gravitational patch gas streams out from the neighbouring voids, towards their encompassing filaments where it shocks, until the cold flows are swallowed by the forming galaxy, advecting their newly acquired AM (Kimm et al. 2011; Tillson et al. 2012). While the gas is streamed out of the walls towards their surrounding filaments it winds up and forms the first generation of galaxies with a spin parallel to the filaments (Aragón-Calvo et al. 2007; Hahn et al. 2007; Paz, Stasyszyn & Padilla 2008; Zhang et al. 2009; Codis et al. 2012; Libeskind et al. 2013, see also Aubert, Pichon & Colombi 2004; Bailin & Steinmetz 2005, for earlier indication of anisotropic inflows). These authors explored the link between DM haloes' spins and the cosmic web to quantify this alignment. They detected a redshift-dependent mass transition $M_{\rm crit}$, varying with the scale (or equivalently with the hierarchical level of the cosmic structure in which the halo is embedded; see Aragón-Calvo & Forrest Yang 2014). Codis et al. (2012) interpreted the correlation in terms of large-scale cosmic flows: high-mass haloes have their spins perpendicular to the filament because they are the results of major mergers (see also Peirani, Mohayaee & de Freitas Pacheco 2004); low-mass haloes are not the products of merger and acquire their mass by accretion, which explains that their spins are parallel to the filament. Danovich et al. (2011) also studied the feeding of massive galaxies at high redshift through cosmic streams using the Horizon-MareNostrum simulation (Ocvirk et al. 2008) and found that galaxies are fed by one dominant stream with a tendency to be fed by three major streams. All these investigations suggest the existence of an additional mechanism affecting first low-mass haloes: mass accretion in an anisotropic, multi-flow environment.

Tempel, Stoica & Saar (2013) and Zhang et al. (2013) have recently found evidence of such alignment in the Sloan Digital Sky Survey (an orthogonality for S0 galaxies and a weak alignment for late-type spirals). The detailed origin of this correlation, while not strictly speaking surprising, as well as its measured dependence on mass, has not yet been fully understood. The spin of the dark halo represents, in essence, the vortical motion of the matter and as such can be expected to reflect the vorticity in the surrounding protogalactic patch of a forming halo. Indeed, to understand this trend, Libeskind et al. (2013) have argued that the local *vorticity* was more relevant than the original tidal field in setting up the direction of dark halo spins. They have explored the link between vorticity in halo environment and the origin of haloes spin and found a strong alignment between both. Vorticity tends to be perpendicular to the axis along which material is collapsing fastest. A natural tell-tale of such process would be a significant large-scale vorticity generation *in* the multiflow regions corresponding to the interior of filaments. Recently, Wang et al. (2014) revisited this description by introducing three invariants of the velocity gradient tensor and concluded that vorticity generation is highly correlated with large-scale structure before and after shell-crossing, in a way which depends on the flow morphology. Vorticity arises only after shell crossing in multi-streaming regions and requires the look inside such regions. Pioneering study of Pichon & Bernardeau (1999) theoretically demonstrated that in the simplest pancake-like multi-stream collapse the level of the vorticity generated is of the order of Hubble constant at the collapse stage at the scale of the thickness of the forming structures. While relying on these theoretical predictions, Codis et al. (2012) speculated that secondary shell-crossing could lead to the formation of vortices aligned with the forming filament. In turn, these vortices would account for the spin of these haloes. There is now indeed ample numerical evidence that the evolution of galaxy morphology is likely to be in part driven by the geometry of the cosmic web, and in particular its vorticity content.

Hence our focus will be in revisiting these findings with an emphasis on *where* (tracing the filaments) and *why* (studying the origin of the vorticity and its orientation) these trends are detected. We will also tentatively explain the origin of the mass transition for halo–spin alignment with the LSS's filaments. This paper aims at revisiting early stages of AM acquisition corresponding to when the cold gas/DM is expelled from neighbouring voids and walls. The main question addressed in this work will be: are there statistical evidence that swirling filaments are responsible for spinning up dark haloes and gaseous discs?

The focus will be specifically exclusively on lower mass haloes ($\mathrm{M}_\odot < 5\,10^{12}$) for which secondary anisotropic infall (originating from the vortex-rich filament within which they form) dominates their AM budget. To that end, we will in particular make use of filament and wall tracers in order to quantify the cosmic web, which is the natural metric for galactic evolution. The virtual data used will be DM and hydrodynamical simulations.

This paper is organized as follows. Section 2 describes the simulations and the estimators implemented in this paper. Section 3 sums up robust statistical results of (i) the orientation of the vorticity relative to the filaments, (ii) the distribution of the vorticity inside the filament and (iii) the alignment of the spin of dark haloes with the vorticity. Section 4 explores qualitatively the origin of this vorticity and uses the link between vorticity and spin to explain the non-monotonic behaviour of spin–filament alignment for haloes with masses lower than $M_{\rm crit}$. Section 5 wraps up and discusses implications. Appendix A finds consistency in alignment with the vorticity of adiabatic/cooling gas. Appendix B illustrates the transition mass in the spin/filament alignment via a simple toy model for the typical vorticity within the caustic. Appendix C studies the effect of persistence and Appendix D the effect of the variation of the smoothing scale on the alignment. Appendix E analyses the orientation of the vorticity with respect to the tidal field eigendirections. Appendix F explains the cleaning of the Friend-of-Friend halo catalogue.

2 DATA SETS AND ESTIMATORS

All the statistical results of Section 3 rely on a set of DM standard Λcold dark matter (ΛCDM) simulations presented in Table 1. These simulations are characterized by the following ΛCDM cosmology: $\Omega_{\rm m} = 0.24$, $\Omega_\Lambda = 0.76$, $n = 0.958$, $H_0 = 73\,{\rm km\,s^{-1}\,Mpc^{-1}}$ and $\sigma_8 = 0.77$ within one standard deviation of 3 year *Wilkinson Microwave Anisotropy Probe* results (Spergel et al. 2003).

We use different box sizes: a 100 h^{-1} Mpc box with an initial mean spatial resolution of 390 h^{-1} kpc (256^3 DM particles) in order to build a statistical sample of haloes and filaments, several 50 h^{-1} Mpc boxes with a mean spatial resolution of 190 h^{-1} kpc (256^3 particles), and a 20 h^{-1} Mpc box with a mean spatial resolution of 39 h^{-1} kpc (512^3 particles). All these simulations were run with GADGET (Springel, Yoshida & White 2001), using a softening length of 1/20th of the mean inter-particle distance. We also use the Horizon-4π simulation, a 2000 h^{-1} Mpc box $\mathcal{S}^{\rm CDM}_{2000}$ with 4096^3 DM particles (Teyssier et al. 2009).

Table 1. The set of simulations used in Sections 3 and 4. The so-called ΛHDM subset corresponds to simulations, the initial condition of which have been smoothed over 2.3 h^{-1} Mpc and 0.23 h^{-1} Mpc. The simulation $\mathcal{S}_{2000}^{\rm CDM}$ corresponds to the post-processing of an HPC simulation which allowed us to identify over 34 million haloes. The velocity field, density field, initial conditions were smoothed with Gaussian filter. In this work, we consider haloes with more than 100 particles.

Name	Type	Box size h^{-1} Mpc	Resolution	$R_{\rm velocity}$ h^{-1} Mpc	$R_{\rm density}$ h^{-1} Mpc	$R_{\rm Lagrangian}$ h^{-1} Mpc	Minimum halo mass $10^{10}{\rm M}_{\odot}$
$\mathcal{S}_{100}^{\rm CDM}$	ΛCDM	100	256^3	0.39	2.3	–	44
$\mathcal{S}_{100}^{\rm HDM}$	ΛHDM	100	256^3	0.39	2.3	2.3	–
$\mathcal{S}_{50}^{\rm CDM}$	ΛCDM	50	$256^3 \times (20)$	0.78	1.2	–	6.2
$\mathcal{S}_{20}^{\rm CDM}$	ΛCDM	20	512^3	0.039	0.23	–	0.044
$\mathcal{S}_{20}^{\rm HDM}$	ΛHDM	20	512^3	0.039	0.23	0.23	–
$\mathcal{S}_{2000}^{\rm CDM}$	ΛCDM	2000	4096^3	–	5	–	77

In addition, the ΛHDM subset corresponds to simulations with initial conditions that have been smoothed with a Gaussian filtering on scales of 2.3 and 0.23 h^{-1} Mpc, respectively, to suppress small-scale modes for the purpose of visualization and interpretation. All simulations but the sets $\mathcal{S}_{50}^{\rm CDM}$, $\mathcal{S}_{2000}^{\rm CDM}$ share the same phases.

All the simulations are studied at redshift $z = 0$. DM haloes are defined thanks to the Friend-of-Friend Algorithm (or FOF; Huchra & Geller 1982), with a linking length of $0.2(L_{\rm box}^3/N_{\rm part})^{1/3}$. In the present work, we only consider haloes with more than 100 particles, which corresponds to a minimum halo mass of 62×10^{10} $\mathrm{M}_{\odot}$ in $\mathcal{S}_{50}^{\rm CDM}$. The spin of a halo is defined as the sum over its particles i: $\sum_i (\boldsymbol{r}_i - \overline{\boldsymbol{r}}) \times (\boldsymbol{v}_i - \overline{\boldsymbol{v}})$ where $\overline{\boldsymbol{r}}$ is the centre of mass of the FOF and $\overline{\boldsymbol{v}}$ its mean velocity. As discussed in Pueblas & Scoccimarro (2009), for the DM simulations we sample optimally the velocity field using a Delaunay tessellation.

The FOF is prone to spuriously link neighbouring structures with tenuous bridges of particles, leading to artificial objects with a very high velocity dispersion, which could eventually bias the measure of the spin and consequently the alignment of the spin and the vorticity. Appendix F investigates the effect of such spurious linkage on vorticity alignments and allows us to conclude that it does not impact the result.

The vorticity of the velocity is then measured from the resampled velocity at each point of the 256^3 grid as the curl of the velocity field $\boldsymbol{\omega} = \nabla \times \boldsymbol{v}$, after Gaussian smoothing of the velocity field with a kernel length of 390 h^{-1} kpc for $\mathcal{S}_{100}^{\rm CDM}$ and $\mathcal{S}_{100}^{\rm HDM}$, a kernel length of 780 h^{-1} kpc for $\mathcal{S}_{50}^{\rm CDM}$ and a kernel length of 39 h^{-1} kpc for $\mathcal{S}_{20}^{\rm CDM}$ and $\mathcal{S}_{20}^{\rm HDM}$. The effect of the smoothing scale on the statistical alignments presented below is investigated in Appendix D. The results do not qualitatively depend on the smoothing scale and the main conclusion remains unchanged, even if the magnitude of the signal varies slightly (but not monotically) according to the scale.

A comparison between vorticity maps in $\mathcal{S}_{20}^{\rm CDM}$ and in $\mathcal{S}_{20}^{\rm HDM}$ is shown in Fig. 1. Vorticity along the normal to the section is plotted in the right panels of this figure. In $\mathcal{S}_{20}^{\rm HDM}$, high frequencies features are suppressed but the low-frequency vorticity remains consistent with that of the more realistic $\mathcal{S}_{20}^{\rm CDM}$. In $\mathcal{S}_{100}^{\rm HDM}$, the smoothing is chosen such that in high-vorticity regions (defined here as being regions where the vorticity is greater than 20 per cent of the maximum vorticity), the mean vorticity is of the order of 90 h km s^{-1} Mpc^{-1}, i.e. it corresponds more or less to one revolution per Hubble time, in agreement with the theoretical predictions of Pichon & Bernardeau (1999). The orders of magnitude are similar in $\mathcal{S}_{100}^{\rm CDM}$, $\mathcal{S}_{20}^{\rm HDM}$ and $\mathcal{S}_{20}^{\rm CDM}$.

The cosmic network is identified with RSEX and DISPERSE, the filament tracing algorithms based on either watershedding (Sousbie, Colombi & Pichon 2009) or persistence (Sousbie 2011; Sousbie, Pichon & Kawahara 2011) without significant difference for the purpose of this investigation. The first method identifies ridges as the boundaries of walls which are themselves the boundaries of voids. The second one identifies them as the 'special' lines connecting topologically robust (filament-like) saddle points to peaks. In this paper, the scale at which the filaments are traced (6 pixels Gaussian for each simulation) corresponds to large enough scales so that we are investigating the flow relative to the LSS (though see Appendix C for variations). Filaments are defined as a set of small segments linking neighbours pixels together. The mean size of the segments is 0.6 pixels, which means 234 h^{-1} kpc in $\mathcal{S}_{100}^{\rm CDM}$.

For comparison with previous studies (e.g. Libeskind et al. 2013), walls are defined according to the density Hessian. Given λ_i the eigenvalues of the Hessian $\mathcal{H} = \partial^2\rho/\partial \mathrm{r}_i \partial \mathrm{r}_j$ where ρ is the density field, with $\lambda_i > \lambda_j$ if $i < j$, walls are identified as being the region of space where $\lambda_1 > \lambda_2 > 0$ and $\lambda_3 < 0$. The normal of a wall is given by the direction of the eigenvector associated with λ_3. To obtain the Hessian, the density field of $\mathcal{S}_{100}^{\rm CDM}$ is smoothed with a Gaussian filter of 1.6 h^{-1}Mpc and differentiation of the density field is performed in Fourier space.

To estimate the number of multi-flow regions within the caustic and their size, for each segment of the skeleton, the vorticity cube is cut with a plane perpendicular to the direction of the filament. The number of multi-flow regions is given by the number of regions of positive and negative projected vorticity along this direction (with a given threshold), counted in a small window centred on the filament. To obtain the size of the regions with a given polarity, the area where the absolute projected vorticity along the normal is greater than 10 per cent of the maximum vorticity is measured, and this area is divided by the number of quadrants. Assuming that these regions are quarter of discs, it yields the corresponding radius. This measure is done in $\mathcal{S}_{100}^{\rm HDM}$.

3 STATISTICAL ALIGNMENTS

Let us first present robust statistical results derived from sets of ΛCDM simulations and ΛHDM simulations for comparison.

3.1 Correlation between vorticity and filaments

The alignment of vorticity with the direction of the filaments is examined in $\mathcal{S}_{100}^{\rm CDM}$ and in $\mathcal{S}_{100}^{\rm HDM}$. The angle μ_1 between the direction

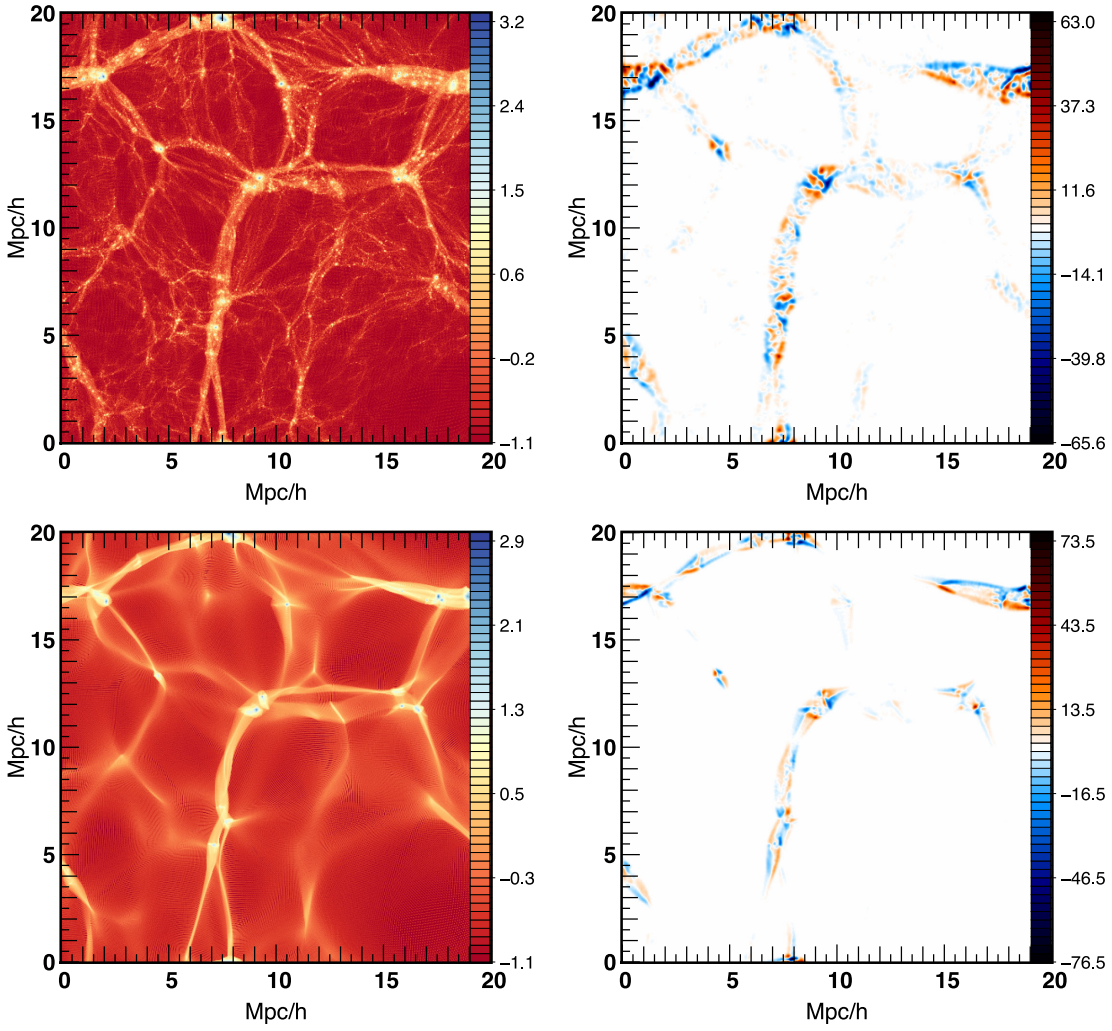


Figure 1. A thin slice ($2\,h^{-1}$ Mpc thickness) of the projected DM density (left panels) and the projected vorticity along the normal to the slice in unit of $h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (right panels). DM density is plotted with a logarithmic scale. Vorticity is computed after smoothing of the velocity field with a Gaussian filter of $160\,h^{-1}$ kpc for this figure only. The geometry of the vorticity closely follows the LSS, but switches polarity across the walls/filaments (recalling that walls appear as filaments and filaments as peaks in such a cross-section). Note also how the vorticity is localized around filaments (the 2D peaks, as exemplified in Fig. 7). The two panels allow for a comparison between a section of $\mathcal{S}_{20}^{\mathrm{CDM}}$ (top) and $\mathcal{S}_{20}^{\mathrm{HDM}}$ (bottom). In $\mathcal{S}_{20}^{\mathrm{HDM}}$, high-frequency modes are suppressed but the low-frequency vorticity is qualitatively consistent with that found in the realistic $\mathcal{S}_{20}^{\mathrm{CDM}}$. On the bottom left panel, the density caustics are quite visible and correspond to the outer edge of the multi-flow region in the bottom right panel.

of the vorticity and the direction of the filament is measured along each segment of the skeleton, and μ_2 between the direction of the vorticity and the direction of the normal of the wall. The probability distribution function (PDF) of the absolute value of the cosine of these angles is shown in Fig. 2. This PDF is normalized for $\cos\mu$ between 0 and 1. A strong detection is achieved. The signal is stronger in $\mathcal{S}_{100}^{\mathrm{HDM}}$ (because of a smoothing of high frequencies) but a clear signal is also detected in $\mathcal{S}_{100}^{\mathrm{CDM}}$. As a check, the alignment between vorticity and shuffled segment directions is then measured: no signal is detected.

In the filaments we find an excess of probability of 20 per cent to have $|\cos\mu_1|$ in [0.5, 1] (that is $0 \leq \mu_1 \leq 60^\circ$) relative to random orientations. In the walls, we find an excess of probability of 45 per cent to have $|\cos\mu_2|$ in [0, 0.5] (that is $60^\circ \leq \mu_2 \leq 90^\circ$) relative to random orientations, which means a strong signal for the vorticity to be aligned with the filament, and perpendicular to the normal of the surrounding wall.

We conclude that in the neighbourhood of filaments, vorticity is preferentially aligned with the filament's axis and perpendicular to the normal of walls. In other words, vorticity tends to be perpendicular to the axis along which material is collapsing fastest. This result is consistent with that of Libeskind et al. (2013), which explored the correlation between vorticity and shear eigenvectors. This correlation is confirmed in Appendix E.

3.2 Geometry of the multi-flow region

Since Section 3.1 showed that vorticity tends to be aligned with the filamentary features of the cosmic web, we are naturally led to focus on the structure of high-vorticity regions. The kinematics of the cross-sections of the filaments is therefore examined, by cutting our simulation with a plane perpendicular to the direction of the filament. We represent in this plane the projected vorticity along the filament. Results are shown in Fig. 3 and can be summarized

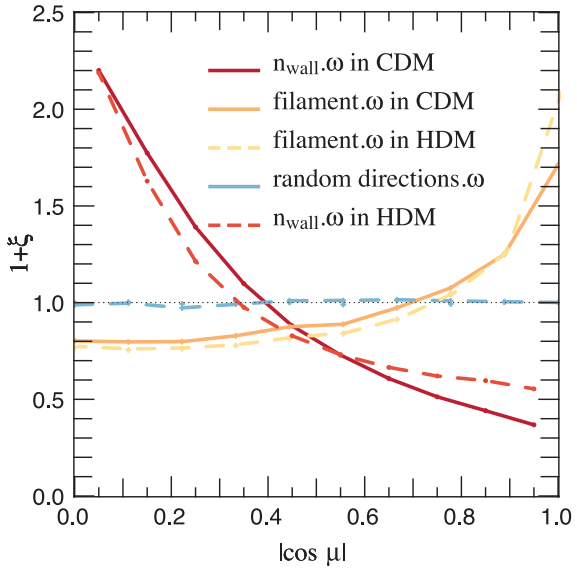


Figure 2. The PDF of $\cos\mu$, the cosine of the angle between the vorticity and the direction of the filament (orange) and the angle between the vorticity and the normal of wall (red) measured in the simulations $\mathcal{S}_{100}^{\rm CDM}$ (solid) and $\mathcal{S}_{100}^{\rm HDM}$ (dashed). The black dotted line corresponds to zero excess probability for reference. The large-scale total vorticity is preferentially aligned with filament axis.

as follows: (i) vorticity is null outside the multi-flow region, and so confined to filaments (and walls in a weaker way) which is consistent with the assumption that cosmic flows are irrotational before shell-crossing; (ii) the cross-sections of the filaments are partitioned into typically quadripolar multi-flow regions (see Fig. 4) where the vorticity is symmetric with respect to the centre of the (density) caustic such that the global vorticity within that caustic is null (as expected); the typical size of each quadrant is of the order of a smoothing scale (as shown in Fig. 4); (iii) high-vorticity resides in the low-density regions of filaments: vorticity is mainly located at the edge of the multi-flow region *on* the caustic (see also Fig. 10); vorticity is in fact typically null *at* the peak of density. (iv) Each quadrant of the multi-flow region is fed by multiple flows, originating from neighbouring walls (see Fig. 9).

3.3 Correlation between vorticity and spin

The alignment of vorticity with filaments on the one hand, and previous results about alignment (or orthogonality) of the low-mass (high-mass) haloes spin with the filament and the shear eigenvectors (Codis et al. 2012; Libeskind et al. 2012) on the other hand, suggests to revisit the alignment of spin with the vorticity (previously examined by Libeskind et al. 2012) and to analyse in depth the correlation between vorticity and AM. The measurement is done by computing the vorticity at the positions of the haloes and the projection, $\cos\theta$, between both normalized vectors. First note that haloes typically stand within one quadrant of the vorticity within filaments and not at the intersection of these quadrants, which is why the spin/vorticity alignment is strong.

The resulting PDF of $\cos\theta$ is displayed in Fig. 5. Here the set of simulations, $\mathcal{S}_{50}^{\rm CDM}$ are used to compute error bars on the correlation between spin and vorticity. The measured correlations are noisier as only a finite number of dark haloes are found within the simulation volume. It was checked that the correlation is not dominated by the intrinsic vorticity of the haloes themselves by computing the alignment between the spin and the vorticity of the field *after extruding* the FOF haloes, which led to no significant difference in the amplitude of the correlation. We find an excess probability of 25 per cent relative to random orientations to have $\cos\theta$ in [0.5, 1] for haloes with $10 \leq \log(M/{\rm M}_\odot) \leq 11$, 55 per cent for $11 \leq \log(M/{\rm M}_\odot) \leq 12$ and 165 per cent for $12 \leq \log(M/{\rm M}_\odot) \leq 13$. Note importantly that the intricate geometry of the multi-flow region (see also Figs 7 and G1) strongly suggests retrospectively that the alignment (including polarity) between the spin of DM haloes and the vorticity of the flow within that region cannot be coincidental.

Fig. 6, which presents PDF of the cosine of the angle between the spin of 43 million dark haloes and the direction of the closest filament identified in the $\mathcal{S}_{2000}^{\rm CDM}$ simulation, displays an interesting

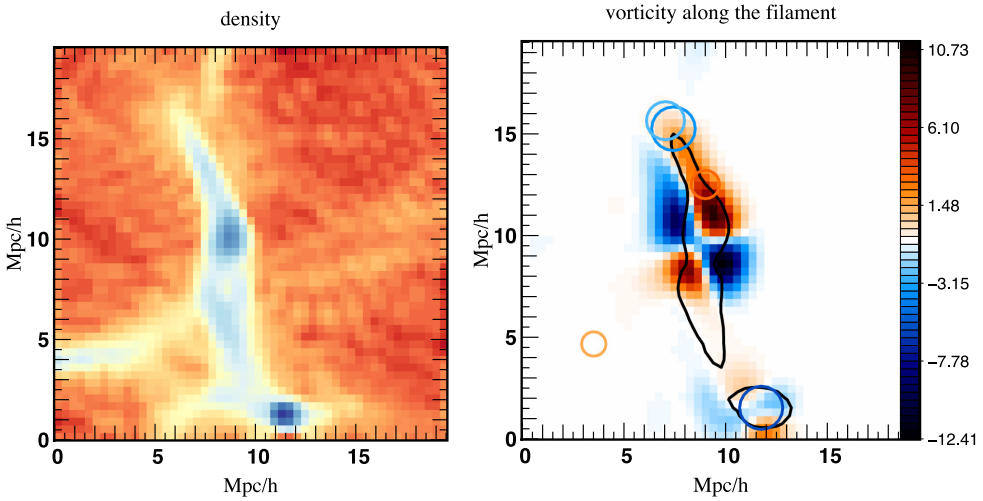


Figure 3. Geometry/kinematics of a typical multi-flow region across a filament. Left: density map of a section perpendicular to a given filament in logarithmic scale. Right: projected vorticity along the filament within that section (towards observer in red and away from the observer in blue) in units of $h\,{\rm km\,s^{-1}\,Mpc^{-1}}$ on which is plotted in dark a contour of the density. Circles are haloes with their corresponding virial radius. The colour of the circles matches to the values of $\cos\theta$ between the haloes spins and the normal of the section, positively oriented towards us. $\mathcal{S}_{100}^{\rm HDM}$ is used here, and for this figure only, vorticity is computed after smoothing the velocity field with a Gaussian filter of $1.6\,h^{-1}$ Mpc.

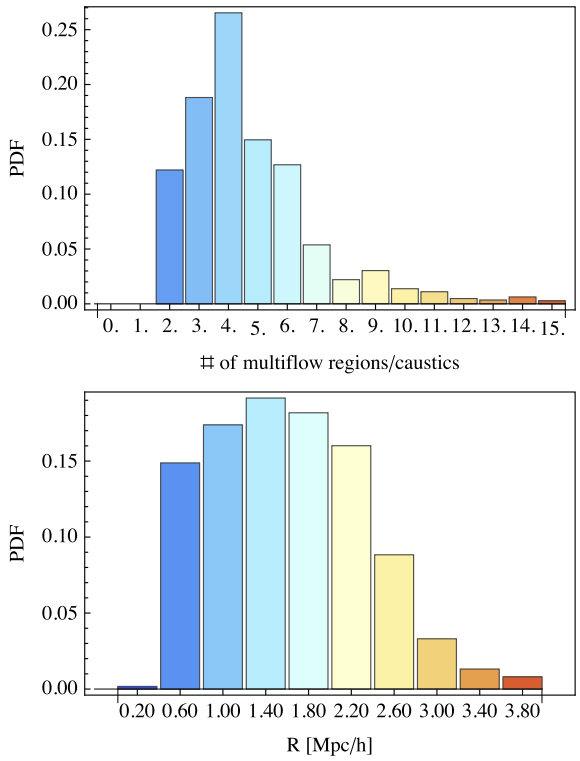


Figure 4. Top: normalized histogram of the number of multi-flow regions with different polarity around a filament measured in the simulation $\mathcal{S}_{100}^{\rm HDM}$. The mean corresponds to $\langle n_{\rm multiflow}\rangle = 4.6$, the median is 4.25. On large scales, the multi-flow region is therefore typically quadrupolar. Bottom: normalized histogram of the size of a region in $\mathcal{S}_{100}^{\rm HDM}$ with a given polarity. The mean size of such region is $\langle R\rangle = 1.6\,h^{-1}$ Mpc, somewhat below the smoothing length of the initial conditions, $R_s = 2.3\,h^{-1}$ Mpc. It was checked on $\mathcal{S}_{20}^{\rm HDM}$ that a similar scaling applies.

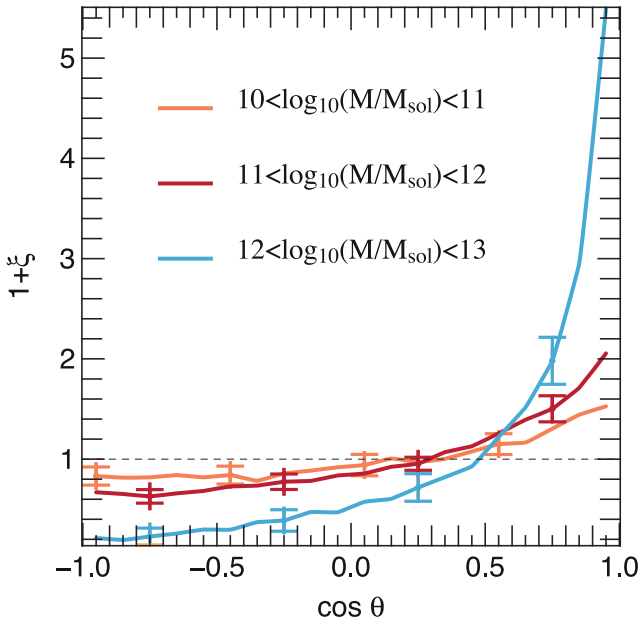


Figure 5. The PDF of the angle between the vorticity and the spin measured in 20 simulations of the $\mathcal{S}_{50}^{\rm CDM}$ set. Haloes are binned by mass as labelled. The displayed error bars are 1σ standard deviation on the mean.

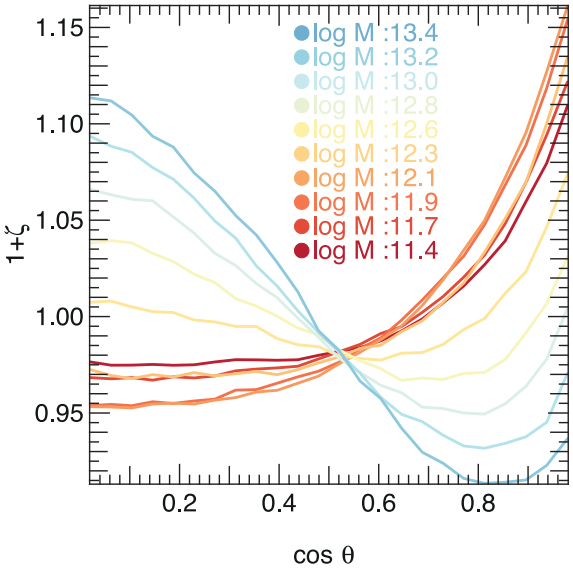


Figure 6. The probability distribution of the cosine of the angle between the spin of dark haloes and the direction of the closest filament as a function of mass in the $\mathcal{S}_{2000}^{\rm CDM}$ simulation. The smoothing length over which filaments are defined is 5 h^{-1} Mpc. This figure extends the result first reported in Codis et al. (2012) to the mass range $\log M/{\rm M}_\odot \sim 11.5$–$12.0$. In this mass range one observes that the probability to have a small angle between the halo's spin and the filament's direction first *increases* (in red) as mass grows to $\log M/{\rm M}_\odot \sim 12.1$, in agreement with the increased spin–vorticity alignment demonstrated in Fig. 5. At larger masses (from orange to blue) the statistical spin–filament alignment quickly decays, with a critical mass (in yellow) corresponding to a transition to predominately orthogonal orientations (in blue) at $\log M_{\rm crit}/{\rm M}_\odot \approx 12.7$ as defined by Codis et al. (2012).

feature at low mass. For the range of mass $\log M/{\rm M}_\odot \sim 11.5$-$12.5$, the actual alignment between the spin and the direction of the filament *increases* with mass, before it becomes abruptly perpendicular around $5 \times 10^{12}\,{\rm M}_\odot$. This is fully consistent with the corresponding increase in vorticity shown in Fig. 5, and will be discussed further in the next section.

4 INTERPRETATION

Let us now turn to the visualization of special purpose simulations, the ΛHDM set, to identify the origin and implications of the measured vorticity of Section 3, and explain the observed mass transition.

4.1 Building up vorticity from LSS flow

Let us first show that density walls are preferentially aligned with zero-vorticity walls.

Fig. 7 displays the vorticity field in the neighbourhood of the main filament of the idealized 'HDM' simulation, $\mathcal{S}_{20}^{\rm HDM}$. The vorticity bundle is clearly coherent on large scales, and aligned with the direction of the filament, strongest within its multi-flow core region, while its essentially quadrupolarity is twisted around it.

Fig. 8 displays the cross-section of the vorticity perpendicular to the main filament shown in Fig. 7. The velocity field lines (in blue) converge towards the local walls (in brown) and are visually in agreement with the vorticity field which is partitioned by these walls.

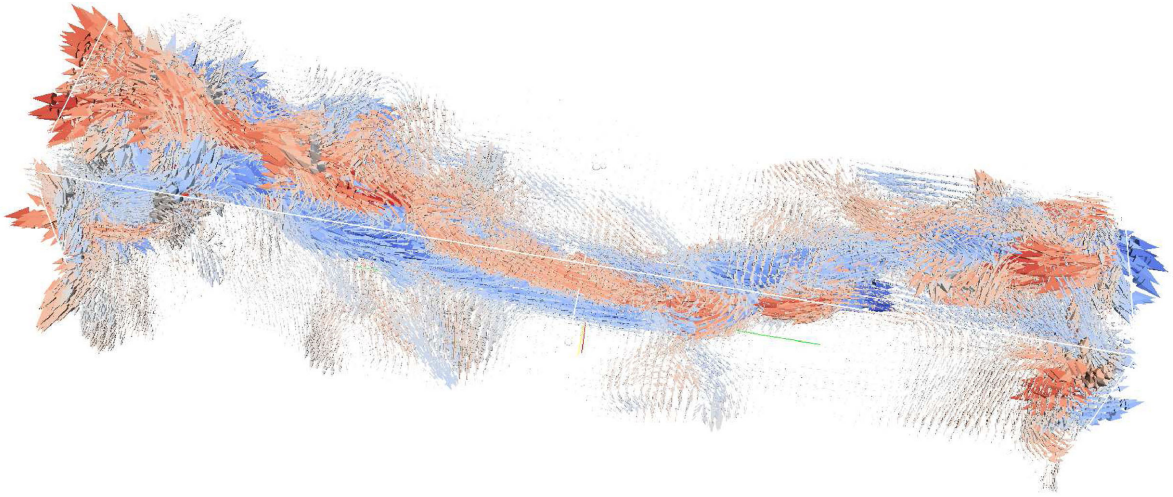


Figure 7. Vorticity field in the neighbourhood of the main filament of the idealized 'HDM' simulation, $\mathcal{S}_{20}^{\rm HDM}$ colour coded through its 'z' component. The vorticity is clearly aligned with the direction of the filament, strongest within its multi-flow core region, while its polarity is twisted around it. Helicity measurements are consistent with the observed level of twisting. We provide animations online at http://www.iap.fr/users/pichon/spin/.

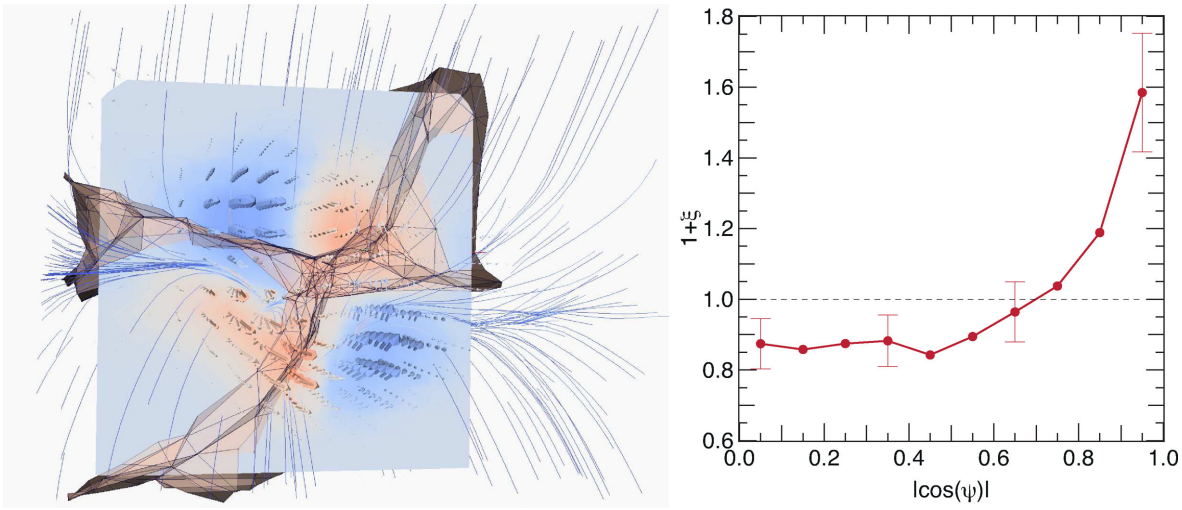


Figure 8. Left: the cross-section of the vorticity perpendicular to the main filament shown in Fig. 7. The colour coding in the section corresponds to the vorticity towards us (in blue) and away from us (in red) as shown by the corresponding arrows. The thin blue lines correspond to velocity field lines. The brown surfaces represent the local walls. The field lines converge towards the local walls and are in agreement with the vorticity field which is partitioned by these walls. Right: the probability distribution as a function of the cosine angle between the normal to the zero vorticity walls and the normal to the density walls, $\cos\psi$, computed on the simulation $\mathcal{S}_{100}^{\rm CDM}$. The simulation is divided into eight 50 h^{-1} Mpc sub-boxes. Density walls are computed using DISPERSE, and the smoothing coefficient of the tessellation is $S = 4$ (see Appendix H). The plotted signal corresponds to the average of the PDFs for the eight sub-boxes. The displayed error bars are 1σ standard deviation on the mean.

This picture is qualitatively consistent with the scenario presented in Codis et al. (2012), as it shows that the filaments are fed via the embedding walls, while the geometry of the flow generates vorticity within their core. This vorticity defines the local environment in which DM haloes form with a spin aligned with that vorticity. The alignment between the contours of minimal vorticity and the density walls which is visually observed in Fig. 8 (left panel) is then quantitatively examined. The probability distribution of the cosine of the angle between the zero vorticity contour and the wall within the caustic is plotted on the right panel of Fig. 8 (see Appendix H for the definition of the zero vorticity contour). An excess of probability of 15 per cent is observed for $\cos\psi$ in [0.5, 1] relative to random distribution, that is for the alignment of the walls with the minimal vorticity contours. This alignment increases with the smoothing of the tessellations, as expected.

4.2 Progenitors of multi-flow region

In a DM (Lagrangian) simulation, it is straightforward to identify the origin of particles within the multi-flow region. Fig. 9 traces back in time DM particles ending up within a quadrant of the multi-flow region. The quadrant is fed by three flows of particles. The flow is irrotational in the initial phase of structure formation until the crossing of three flows in the vicinity of the filaments generates shear and vorticity close to the caustic.

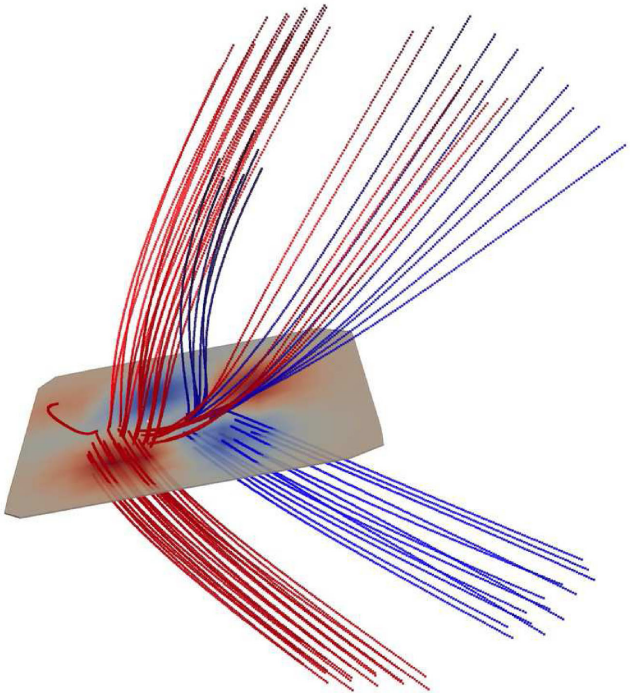


Figure 9. Left: individual DM particle trajectories ending in a given quadrant of the vorticity multi-flow region. In blue are particles ending in a region of positive projected vorticity along the filament, and in red are the particles ending in the negative vorticity region. The quadrant is fed by at least three flows of particles (see also the inset in Fig. 11, which represents qualitatively the theoretical expectation of the starting points of these bundles in the Zel'dovich approximation.). The $\mathcal{S}_{100}^{\rm HDM}$ simulation is used for this figure.

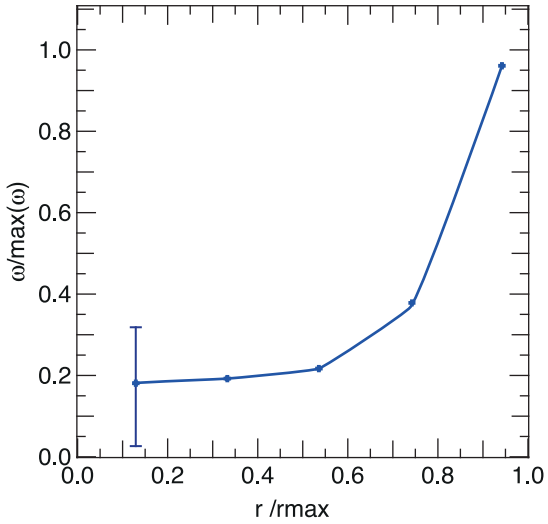


Figure 10. Azimuthal average of the radial profile of the vorticity. The profile is obtained by averaging on the sections of a complete filament (each section is associated with a filament segment, to which the section is perpendicular). Vorticity is clearly larger towards the caustic, and would theoretically become singular (as $1/\sqrt{1-r/r_{\rm max}}$) *at* the caustic for a Zel'dovich mapping, as shown in Pichon & Bernardeau (1999). Here the profile is convolved by shape variations from one caustic to another and by the azimuthal average. The indicative error bar was computed as the average over a larger stack.

Note that the sharp rise near the edge of the multi-flow region at the caustic is qualitatively consistent with catastrophe theory (Arnold 1992), and is directly related to the prediction of Pichon & Bernardeau (1999). Fig. 10 illustrates this fact. To obtain this profile, a filament is cut in slices, corresponding to filament segments: each slice corresponds to a plane perpendicular to the direction of the segment. Local vorticity is measured within that plane and stacked. The amount of vorticity is greater near the caustic. These results are qualitatively consistent with the above-mentioned theoretical predictions which characterize the size and shape of the multi-flow regions after first shell crossing, and estimate their vorticity content as a function of cosmic time.

In short, having looked in detail at the set of (Lagrangian-) smoothed simulations allows us to conclude that streaming motion of DM *away* from minima and wall-saddle points of the field, and *along* the walls of the density field is responsible for generating the multi-flow region in which vorticity arises. In turn, this vorticity defines the environment in which lower mass haloes collapse. Such haloes inherit their spin from this environment, as quantified by Fig. 5.

4.3 Mass transition for spin–filament alignment

Up to now, we have not considered the mass of the forming halo within the multi-flow region. The assumption has been that the Lagrangian extension of the progenitor of the dark halo was small compared to the antecedents of a given vorticity quadrant, so that the collapse occurs within a quadrant of a given polarization, and leads to the formation of haloes with a spin parallel to that vorticity. For more massive objects (of the order of the transition mass), we can anticipate that their progenitor patch overlaps future vorticity quadrants of opposite polarity, hence that they will mostly cancel the component of their vorticity aligned with the filament as they form. The above-mentioned observed transition mass between aligned and anti-aligned spins relative to filaments would then typically correspond to the mass associated with the width of the quadrant of each caustics. In fact, as argued in Pichon & Bernardeau (1999, fig. 7) and shown in Fig. 10, the vorticity within the multi-flow region is mostly distributed near the caustic, on the outer edge of the multi-flow region. It is therefore expected that, as the size of the collapsed halo increases, but remains *below* that of the quadrant, its vorticity should increase (as it collects more and more coherent rotating flow as secondary inflow), as shown in Fig. 5. As it reaches sizes *above* that of the quadrant, it should start to diminish significantly[1] (see also Fig. 11 and Appendix B where this transition is illustrated with the help of a toy model).

Let us turn back specifically to Fig. 6. For the range of mass $\log M/{\rm M}_\odot \approx 11.4$–$12.1$, the alignment between the spin and the direction of the filament *increases* with mass peaking at $M_{\rm max} \approx 10^{12}\,{\rm M}_\odot$, before it rapidly decreases and changes to preferably perpendicular one for $\log M > \log M_{\rm crit} \approx 12.7$, i.e. $M_{\rm crit} \approx 5 \times 10^{12}\,{\rm M}_\odot$. This is fully consistent with the corresponding increase in vorticity shown in Fig. 5.

The characteristic masses can be roughly understood by conjecturing that the highest alignment occurs for the haloes which are of the size of vortices in the caustic regions that *just* undergo collapse. The measured caustic structure depends on the chosen smoothing scale, so a recently formed filament corresponds to the vortex that

[1] In fact, while investigating the statistics of the vorticity within spherical shells, Pichon & Bernardeau (1999) showed that if we consider spheres of size above one quadrant of the multi-flow regions, the total vorticity within that sphere drops significantly.

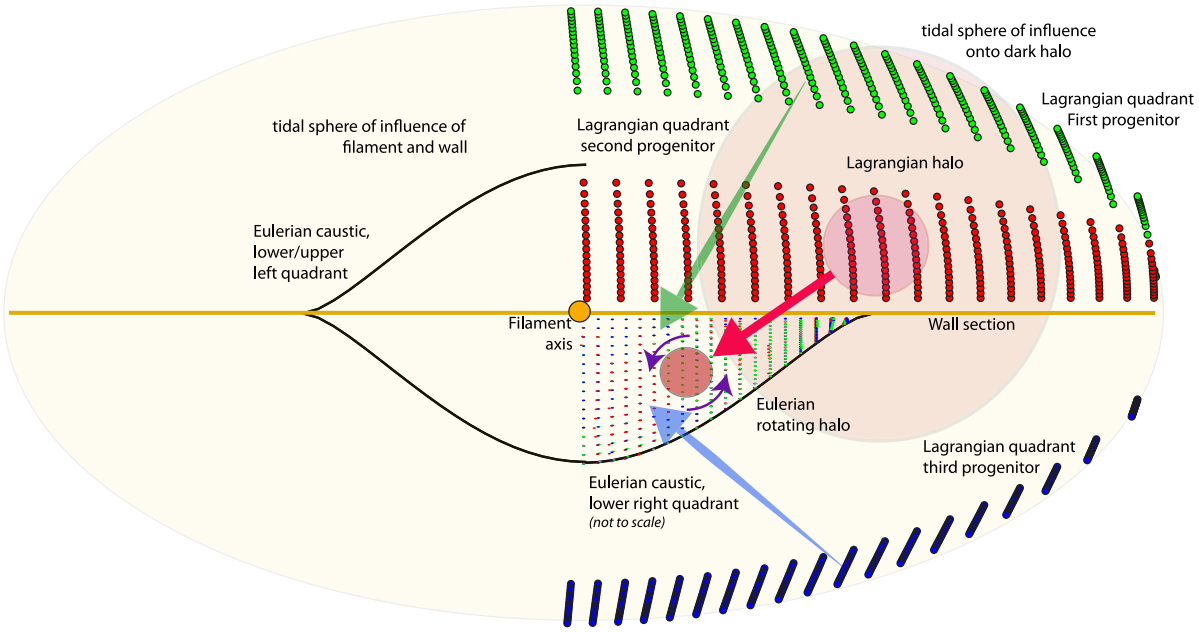


Figure 11. Sketch of the dynamics of a low-mass halo formation and spin-up within a wall near a filament, which are perpendicular to the plane of the image (in yellow). The tidal sphere of influence of this structure is represented by the pale yellow ellipse. The three bundles of large dots (in green, red, and blue) represent Lagrangian points (at high redshift) which image, after shell crossing, will end up sampling regularly the lower right quadrant of the Eulerian multi-flow region; the three progenitor bundles are computed here in the Zel'dovich approximation (see Pichon & Bernardeau 1999, for details; note that this Eulerian quadrant is not up to scale). Each pair of dots (one large, one small) represents the same DM particle in the initial condition and final condition. In black, 3/4 of the Eulerian caustic. In light, resp. dark pink, the locus of the Lagrangian and Eulerian position of the halo, which has moved by a distance displayed by the red arrow, and spun up following the purple arrows while entering the quadrant. The blue and green arrows represent the path of fly-by DM particles originating from the other two bundles, which will contribute to torquing up the halo (following Codis et al. 2012). Given the geometry of the flow imposed by the wall and swirling filament, the spin of the DM halo will necessarily be parallel to the direction of the filament *and* to the vorticity in that quadrant. In the language of TTT, the tidal field imposed on to the Lagrangian patch of the halo (very light pink, corresponding to secondary infall) should be evaluated subject to the constraint that the halo will *move* into the *anisotropic multi-flow* region (each emphasis imposing a constraint of its own); these constraints will in turn impose that the corresponding spin-up will be aligned with the vortex.

shows a basic four quadrant structure and, following Pichon & Bernardeau (1999), which has vorticity close to the Hubble value H. From our simulations, the typical Lagrangian radius of such vortex is $\approx 1.5\,h^{-1}$ Mpc, which if taken as the top-hat scale gives a mass estimate $M_{\rm max} \approx 1.5 \times 10^{12}\,{\rm M}_{\odot}$ for the mass of haloes with maximum spin/filament alignment. The transition to misalignment will happen at $M_{\rm crit} \approx 8 \times M_{\rm max}$ where the whole width of the filament is encompassed. Of course the quantitative accuracy of such argument should not be over-emphasized. For instance, if we took the Lagrangian radius of the vortex to be $1.3\,h^{-1}$ Mpc, we would get $M_{\rm max} = 10^{12}\,{\rm M}_{\odot}$, which would fit the transition of Fig. 6 even closer.

5 DISCUSSIONS AND CONCLUSIONS

Let us reframe the findings of Section 3 and 4 in the context of recent published results in this field before concluding.

5.1 Discussion

Libeskind et al. (2013)'s description of AM acquisition occurs in two stages (first through TTT, and then through the curl of the embedding velocity field). Results of Sections 3 and 4 seem consistent with this. In particular the alignment of vorticity with the eigenvectors of the tidal field is confirmed in Appendix E. The connection between Pichon et al. (2011) and this paper is the following: in the former, it was shown that the spin up of dark haloes proceeded in stages: a given collapsing halo would first acquire some specific AM following TTT, at turn around freezing its amplitude at the TTT expected value; in a secondary stage (see their fig. 9), it would spin up again as it acquires specific AM from secondary infall coming from the larger scale distribution of matter which collapses at the next stage of hierarchical clustering. For relatively isolated massive haloes that form from statistically rare density enhancements as studied in Pichon et al. (2011), this secondary collapse just leads to a virialized halo of increased mass. The (Eulerian) emphasis of the current paper and of Codis et al. (2012) (see also Danovich et al. 2011) is to note that for less massive and less rare haloes, forming in large-scale filamentary regions, this secondary infall, coming late from the turn-around of the encompassing filamentary structure, is arriving along marked preferred directions and is typically multi-flow and vorticity rich. Given that the shell crossing occurring during the later formation of that embedding filament generates vorticity predominantly aligned with the filament, this secondary infall will contribute extra spin up along the filament direction. Hence the global geometry of the inflow is consistent with a spin acquisition for haloes induced by the large-scale dynamics within the cosmic web, and in particular its multi-flow vortices. This scenario may only be reconciled with the standard (Lagrangian) tidal torque theory if the latter is augmented so as to account for the larger scale anisotropic environment of walls and filaments responsible for secondary infall (see Pichon et al. 2014).

Let us sketch the basis of such calculation. The geometry of the setting is shown in Fig. 11. For Gaussian random fields, one can

Table 2. The median and mean cosine values for the set of studied alignments. In parenthesis are the expected values for random distributions.

Definition	Name	Mean	Median
Alignment between vorticity and Cosmic Web			
DM: vorticity/filaments		0.58 (0.5)	0.62
Hydro: vorticity/filaments	$\lvert\cos\mu\rvert$	0.58 (0.5)	0.63
DM: vorticity/walls		0.34 (0.5)	0.27
Alignment between vorticity and haloes spin			
$10 \leq \log(M/\mathrm{M}_\odot) \leq 11$		0.09 (0.0)	0.14
$10 \leq \log(M/\mathrm{M}_\odot) \leq 12$	$\cos\theta$	0.19 (0.0)	0.29
$12 \leq \log(M/\mathrm{M}_\odot) \leq 13$		0.53 (0.0)	0.72
Alignment between density walls and 0-vorticity walls			
	$\lvert\cos\psi\rvert$	0.54 (0.5)	0.56
Alignment between vorticity and tidal tensor eigenvectors			
Vorticity/e1		0.62 (0.5)	0.69
Vorticity/e2	$\lvert\cos\gamma\rvert$	0.48 (0.5)	0.47
Vorticity/e3		0.31 (0.5)	0.23

compute the most likely tidal field and inertia tensor at a given Lagrangian peak, subject to a Zel'dovich boost which will translate that peak near to a filament at some distance $\boldsymbol{r}$; this distance corresponds to the time during which the nearby filament has shell-crossed multiplied by the original velocity. In turn, the condition of shell crossing can be expressed as constraints on the eigenvalues of the shear tensor. We can anticipate that the pre-existing Lagrangian correlation between the tidal field of the halo-to-be, on the one hand, and the Hessian of the filament-to-be, on the other hand, imposes some alignment between the direction of the filament (along the first eigenvector of the Hessian) and the spin of the collapsing halo (as set by the corresponding tidal tensor). If the critical condition that the filament is embedded into a given wall is added, the axial symmetry of the problem will be broken, and the inertia and tidal tensor (which are sensitive to different scales) will end up misaligned, reflecting this anisotropy. In this context, the observed spin (and importantly its polarity) will correlate with the polarity of the vorticity quadrant the halo ends up into after translation. The upshot is that in Fig. 11; the lighter pink sphere will 'know' about the green dots given these constraints. This supplementary requirement is imposed by the fact that the correlation between spin and vorticity keeps tract of the direction of both vectors, as shown in Fig. 5. It appears from this sketch that, as the Lagrangian patch of the protohalo becomes of the order of the typical Lagrangian size of the quadrant, the alignment will increase, and as it becomes larger, it will fade (see Appendix B for an illustration of this transition). Note finally that this 'one slice perpendicular to the filament axis' picture cannot address the process of spin flipping to a perpendicular direction to the filament via mergers, as this is a longitudinal process. This is also the topic of Pichon et al. (2014) which complements the Eulerian view presented here.

5.2 Conclusions

Using large-scale cosmological simulations of structure formation, we have analysed the kinematic properties of the velocity flows relative to the cosmic web. Our findings are the following.

(i) The vorticity in large-scale structures on scales of 0.39 h^{-1} Mpc and above is confined to, and aligned with, its filaments with an excess of probability of 20 per cent relative to random orientations, and perpendicular to the normal of the dominant walls at a similar level. This is consistent with the corresponding direction of the eigenvectors of the tidal field (and is expected given that the potential is a smoothed version of the density field).

(ii) At these scales, the cross-sections of these filaments are typically partitioned into quadripolar caustics, with opposite vorticity parallel to their filament, arising from multiple flows originating from neighbouring walls, as would secondary shell crossing along these walls imply. The radial vorticity profile within the multi-flow region displays a sharp rise near the caustic, a qualitatively expected feature of catastrophe theory.

(iii) The spins of embedded haloes within these filaments are consistently aligned with the vorticity of their host vorticity quadrant at a level of 165 per cent. The progenitor of lighter haloes within the multi-flow region can be traced back to three flows or more originating from the neighbouring walls, and form within the filament.

(iv) Appendix A shows that for adiabatic/cooling hydrodynamical simulations within the DM caustics, the gas and the DM share the same vorticity orientation on large scales. High-resolution cooling runs show that the small-scale structure of the velocity flow around forming galaxies does not destroy this larger scale coherence.

(v) The mass transition for spin–filament alignment is set by the size of sub-caustics with a given polarization (see Appendix B). The alignment is strongest for Lagrangian patch commensurable with the sub-caustic as vorticity is strongest on the edge of the multi-flow region. Once the collapsed halo has a size larger than any such sub-caustic, it cancels out most of the vorticity within the caustics.

The focus of this paper was in explaining the 'where': pinning down the locus of vorticity and describing the geometry of multi-flow infall towards filaments; and the 'how': explaining its origin via shell crossing. It also provided an explanation for the origin of the mass transition for spin alignment. All measured alignments are summarized in Table 2.

Improvements beyond the scope of this paper include (i) developing the sketched anisotropic (filamentary) peak-background-split theory of spin acquisition; (ii) quantifying the curvilinear evolution of the vorticity (orientation and amplitude) as a function of distance to the critical points of the cosmic web and predicting the spin flip for high masses; (iii) quantifying the helicoidal nature of gas infall

on galactic scales; (iv) connecting the findings of this paper to the actual process of *galactic* alignment.

In turn, this should allow astronomers to shed light on the following problems: how and when was the present Hubble sequence of galaxies established? How much of the dynamical evolution of galaxies is driven by environment? What physical processes transforming galaxies dominate morphology: galaxy interactions and mergers, external accretion and outflows, secular evolution? What is their respective roles in shaping discs, bulges or spheroids? Is it the same process at low and high redshift? These are addressed in part in the companion paper, Dubois et al. (2014), which shows in particular using state-of-the-art hydrodynamical simulations with AGN/SN feedback that at high redshifts the large vorticity of the gas flow is correlated with the direction of the spin of *galaxies* (their fig. 12).

ACKNOWLEDGEMENTS

We thank M. Haehnelt, James Binney and D. Lynden-Bell for useful comments during the course of this work. JD and AS's research is supported by Adrian Beecroft, the Oxford Martin School and STFC. Let us thank D. Munro for freely distributing his YORICK programming language and OPENGL interface (available at http://yorick.sourceforge.net). Some visualizations made use of the software PARAVIEW http://www.paraview.org. This work is partially supported by the Spin(e) grants ANR-13-BS05-0005 of the French *Agence Nationale de la Recherche* and by the ILP LABEX (under reference ANR-10-LABX-63 and ANR-11-IDEX-0004-02). CP thanks the PEPS 'Physique théorique et ses interfaces' for funding, the institute of Astronomy for a Sacker visiting fellowship and the KITP for hospitality via the National Science Foundation under Grant No. NSF PHY11-25915.

REFERENCES

Aragón-Calvo M. A., Forrest Yang L., 2014, MNRAS, 440, L46
Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007, ApJ, 655, L5
Arnold V. I., 1992, Catastrophe Theory/V.I. Arnold [translated from the Russian by G.S. Wassermann; based on a translation by R.K. Thomas, 3rd rev. and expanded edn.]. Springer-Verlag, Berlin, New York
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bailin J., Steinmetz M., 2005, ApJ, 627
Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215
Birnboim Y., Dekel A., 2003, MNRAS, 345, 349
Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320
Danovich M., Dekel A., Hahn O., Teyssier R., 2011, MNRAS, 422, 1732
Doroshkevich A. G., 1970, Astrofizika, 6, 581
Dubois Y., Pichon C., Haehnelt M., Kimm T., Slyz A., Devriendt J., Pogosyan D., 2012, MNRAS, 423, 3616
Dubois Y., Pichon C., Welker C., Le Borgne D., Devriendt J., Laigle C., Codis S., Pogosyan D., 2014, MNRAS, 444, 1453
Haardt F., Madau P., 1996, ApJ, 461, 20
Hahn O., Carollo C. M., Porciani C., Dekel A., 2007, MNRAS, 381, 41
Hoyle F., 1949, Problems of Cosmical Aerodynamics. Central Air Documents, Office, Dayton, OH
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Katz N., Keres D., Dave R., Weinberg D. H., 2003, in Rosenberg J. L., Putman M. E., eds, The IGM/Galaxy Connection. The Distribution of Baryons at z=0. Astrophysics and Space Science Library, Vol. 281, How Do Galaxies Get Their Gas? p. 185
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kimm T., Devriendt J., Slyz A., Pichon C., Kassin S. A., Dubois Y., 2011, preprint (arXiv:1106.0538)
Libeskind N. I., Hoffman Y., Knebe A., Steinmetz M., Gottlöber S., Metuki O., Yepes G., 2012, MNRAS, 421, L137
Libeskind N. I., Hoffman Y., Forero-Romero J., Gottlöber S., Knebe A., Steinmetz M., Klypin A., 2013, MNRAS, 428, 2489
Libeskind N. I., Hoffman Y., Steinmetz M., Gottlöber S., Knebe A., Hess S., 2013, ApJ, 766, L15
Ocvirk P., Pichon C., Teyssier R., 2008, MNRAS, 390, 1326
Paz D. J., Stasyszyn F., Padilla N. D., 2008, MNRAS, 389, 1127p
Peebles P. J. E., 1969, ApJ, 155, 393
Peirani S., Mohayaee R., de Freitas Pacheco J. A., 2004, MNRAS, 348, 921
Pichon C., Bernardeau F., 1999, A&A, 343, 663
Pichon C., Pogosyan D., Kimm T., Slyz A., Devriendt J., Dubois Y., 2011, MNRAS, 418, 2493
Pichon C., Codis S., Pogosyan D., Dubois Y., Desjacques V., Devriendt J., 2014, preprint (arXiv:1409.2608)
Press W. H., Schechter P., 1974, ApJ, 187, 425
Prieto J., Jimenez R., Haiman Z., 2013, MNRAS, 436, 2301
Pueblas S., Scoccimarro R., 2009, Phys. Rev. D, 80, 043504
Schaefer B. M., 2009, Int. J. Modern Phys. D, 18, 173
Sousbie T., 2011, MNRAS, 414, 350
Sousbie T., Colombi S., Pichon C., 2009, MNRAS, 393, 457
Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384
Spergel D. N. et al., 2003, ApJS, 148, 175
Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79
Stewart K. R., Brooks A. M., Bullock J. S., Maller A. H., Diemand J., Wadsley J., Moustakas L. A., 2013, ApJ, 769, 74
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Tempel E., Stoica R. S., Saar E., 2013, MNRAS, 428, 1827
Teyssier R., 2002, A&A, 385, 337
Teyssier R. et al., 2009, A&A, 497, 335
Tillson H., Devriendt J., Slyz A., Miller L., Pichon C., 2012, preprint (arXiv:1211.3124)
Wang X., Szalay A., Aragón-Calvo M. A., Neyrinck M. C., Eyink G. L., 2014, ApJ, 793, 58
White S. D. M., 1984, ApJ, 286, 38
Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747
Zhang Y., Yang X., Wang H., Wang L., Mo H. J., van den Bosch F. C., 2013, ApJ, 779, 160

APPENDIX A: THE VORTICITY OF THE GAS

We use three hydrodynamical simulations $\mathcal{S}_{100}^{\rm HA}$, $\mathcal{S}_{100}^{\rm HC}$ and $\mathcal{S}_{20}^{\rm cool}(0.7)$ (see also Table A1), carried out with the Eulerian hydrodynamic code RAMSES (Teyssier 2002), which uses an Adaptative Mesh Refinement (AMR) technique. For these hydrodynamical runs, the evolution of the gas is followed using a second-order unsplit Godunov scheme for the Euler equations. The HLLC Riemann solver with a first-order MinMod Total Variation Diminishing scheme to

Table A1. The set of hydrodynamical simulations used in Appendix A. The hydro runs come in two categories: adiabatic and cooling, including one high-resolution run which was stopped at redshift 0.7.

Name	Type	Box size h^{-1} Mpc	Resolution
$\mathcal{S}_{100}^{\rm HA}$	ΛHDM adiabatic	100	256^3
$\mathcal{S}_{100}^{\rm HC}$	ΛHDM cool	100	256^3
$\mathcal{S}_{20}^{\rm cool}(0.7)$	ΛCDM cool	20	1024^3

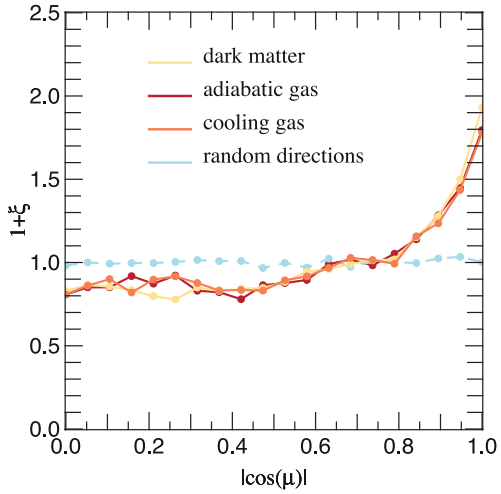


Figure A1. The probability distribution of the cosine of the angle between the vorticity in the smoothed DM and hydrodynamical simulations and the direction of the filament (solid line). The same measure is done for random directions of u (dashed blue line), plotted for the DM in $\mathcal{S}_{100}^{\rm HDM}$ and for the gas in $\mathcal{S}_{100}^{\rm HA}$ (adiabatic gas) and $\mathcal{S}_{100}^{\rm HC}$ (cooling run). We find an excess of probability for $|\cos\mu|$ in [0.5, 1] relative to random orientations, and three profiles are very similar, which shows that large-scale modes dominate.

reconstruct the interpolated variables from their cell-centred values is used to compute fluxes at cell interfaces. Collisionless particles (DM and star particles) are evolved using a particle-mesh solver with a Cloud-In-Cell interpolation. The initial mesh is refined up to $\Delta x = 1.7$ kpc according to a quasi-Lagrangian criterion: if the number of DM particles in a cell is more than eight, or if the total baryonic mass in a cell is eight times the initial DM mass resolution.

For the cooling runs $\mathcal{S}_{100}^{\rm cool}$, and $\mathcal{S}_{20}^{\rm cool}$, gas is allowed to cool by H and He cooling with an eventual contribution from metals using a Sutherland & Dopita (1993) model down to 10^4 K. Heating from a uniform UV background takes place after redshift $z_{\rm reion} = 10$ following Haardt & Madau (1996).

On large scales (as probed by the smoothed sets of simulations) the vorticity of gas shows the same correlations with the filaments as DM does. Fig. A1 displays the probability distribution of the cosine of the angle between the vorticity and the direction of the filament for the DM field (in red), the adiabatic gas (in blue) cooling run (in yellow). These three simulations quantitatively show the same preference for their vorticity to be aligned with the filamentary structure. In a nutshell, differences between the adiabatic and the cooling run only appear on kpc scales, so that on large scales, the DM, adiabatic and cooling runs have the same velocity field structure.

Fig. C1 displays the probability distribution of the cosine of the angle between the vorticity and the direction of the skeleton for a range of redshifts. The correlation between the direction of the filament and the vorticity is significant. As expected, this correlation decreases with cosmic time (at a fixed smoothing scale). Appendix C investigates the evolution of this correlation as a function of the skeleton's persistence. As long as we consider large enough scales, the alignment pervades and is consistent with that of the DM. On smaller scales, the gas is dense enough to allow cooling to operate and re-structure the velocity flow. Notwithstanding, these smaller scale structures do not affect the larger scale correlation between vorticity and the direction of the filaments.

APPENDIX B: TOY MODEL FOR HALO SPIN

Can a model based on a vorticity field in qualitative agreement with what was found in the simulation explain why it should lead the observed evolution of spin alignment with mass? Let us qualitatively illustrate with a simple toy model this mass transition for the spin–filament alignment. In this toy problem, we consider an isolated infinite filament aligned along $\boldsymbol{e}_z$. We define the corresponding idealized vorticity field as

$$\boldsymbol{\Omega}(\boldsymbol{r},\theta) = C\epsilon\,\sin(2\theta)\frac{1}{(\epsilon^2+(r-R)^2)}\,\boldsymbol{e}_z\,,$$

with C a constant, R the radius of the caustic and ϵ a small number. The vorticity thus defined is largest along the caustic, point reflection symmetric and tends rapidly to 0 outside the caustic. Should ϵ tend to 0, vorticity would become singular on the edge of the caustic ($r \to R$). The map of the vorticity is displayed in Fig. B1 (top left panel).

By application of the Helmoltz–Hodge theorem, we find that the curl component of the velocity field consistent with that vorticity (i.e. such that $\boldsymbol{\Omega} = \nabla \times \boldsymbol{v}$) obeys

$$\boldsymbol{v}(\boldsymbol{r},\theta) = \frac{1}{4\pi}\int_{\mathcal{V}} \nabla\times\boldsymbol{\Omega}(\boldsymbol{r}',\theta')\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}{\rm d}\mathcal{V}\,. \qquad \text{(B1)}$$

We assume here that the shear part of the curl free component of the velocity flow is smaller on scales comparable to the halo. We now consider a spherical halo of radius $r_{\rm h}$ embedded in one of the four quadrants of the caustic, centred on $C_{\rm h}(x_{\rm h}, y_{\rm h})$ with $x_{\rm h}^2 + y_{\rm h}^2 \leq R$. From equation (B1) we can simply compute its AM $\mathcal{J}(r_{\rm h}, x_{\rm h}, y_{\rm h})$, and look at the variation of $\mathcal{J}$ as a function of its position at fixed radius, or as a function of its radius at fixed position. Fig. B1 (bottom panel) shows the magnitude of the AM along the z-axis for a halo centred on $C_{\rm h}(0.5\,R/\sqrt{2},0.5\,R/\sqrt{2})$ as a function of the radius. We observe that the alignment increases until the size of the halo encompasses the whole quadrant. At a given radius, the position of the halo which maximizes the AM is the one for which the edge of the halo coincides with the edge of the caustic, since the vorticity peaks close to the caustic.

APPENDIX C: PERSISTENCE EFFECTS

Given the characteristics of ΛCDM hierarchical clustering, one can anticipate that the process described in the main text occurs on several nested scales at various epochs – and arguably on various scales at the same epoch. The scenario we propose for the origin of vorticity and spin alignment is, like the signal itself, relative to the linear scale involved in defining the filaments and as such multi-scale. Indeed in the main text, the two sets of simulations, $\mathcal{S}^{\Lambda\rm CDM}$ and $\mathcal{S}^{\Lambda\rm HDM}$, allowed us to probe different scales of the vorticity field. The induced multi-scale anisotropic flow also transpires in the scaling of the spin flipping transition mass with smoothing presented in appendix of Codis et al. (2012). It will hold as long as filaments are well defined in order to drive the local cosmic flow.

Let us now briefly explore the effects of probing different scales of the LSS via the skeleton level of persistence. Fig. C1 shows the excess alignment probability as a function of the cosine of the angle between the vorticity and the filaments as a function of the persistence level for a range of values. The alignment is strongest

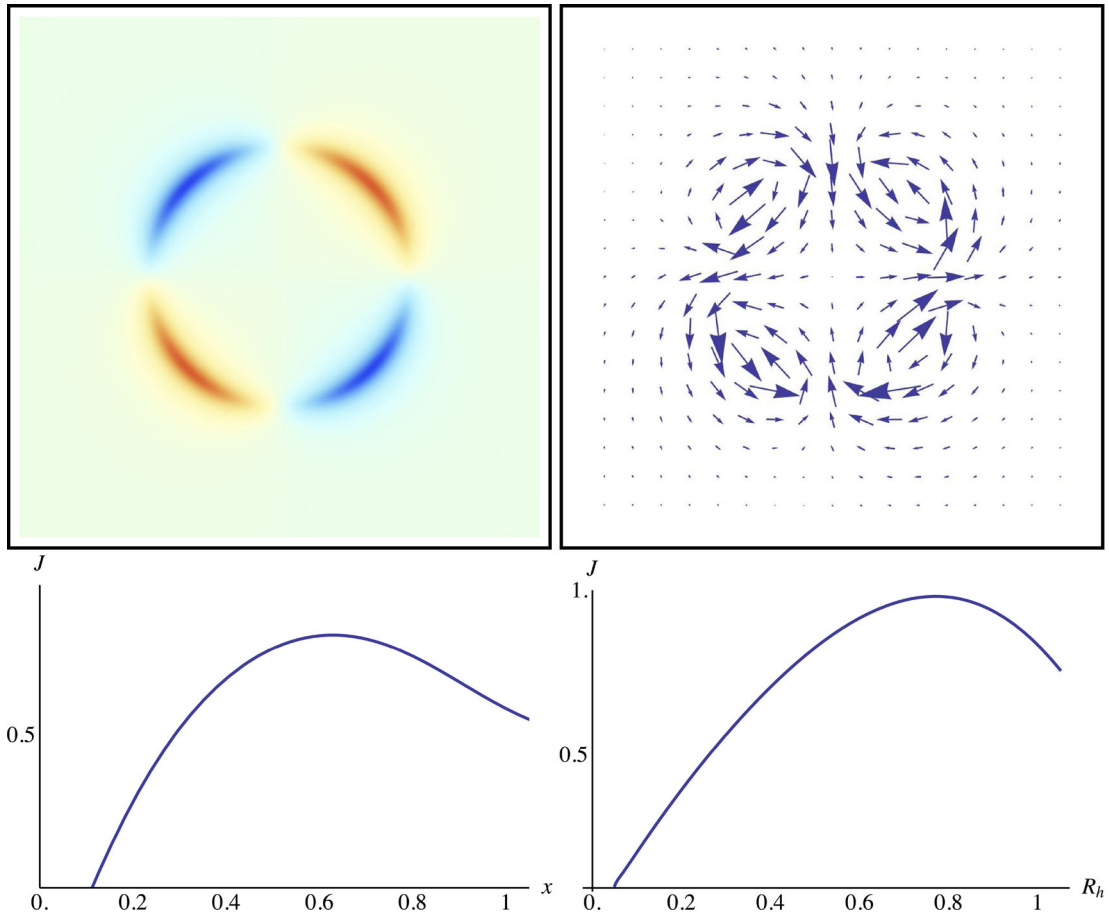


Figure B1. Top: maps of the vorticity (left) projected along $\boldsymbol{e}_z$ and the associated curl component of the velocity (right) computed from the vorticity by application of the Helmoltz–Hodge theorem. Bottom: magnitude of the AM along $\boldsymbol{e}_z$ for a halo embedded in one of the four quadrants. We first consider how it varies as a function of the radius of the halo (left), the position of this latter being fixed $(x_\mathrm{h},y_\mathrm{h}) = (0.5/\sqrt{2}, 0.5/\sqrt{2})$. The alignment of the AM of the halo with the vorticity increases until the halo size becomes comparable to that of the vorticity quadrant. We study then the magnitude of the AM along $\boldsymbol{e}_z$ at fixed radius ($R_\mathrm{h} = 0.3$) as a function of the position x along the diagonal (right). In this case, the alignment increases up to the point where the halo boundary coincides with the vorticity caustic, with the halo still being fully contained within the vorticity quadrant $(x_\mathrm{h},y_\mathrm{h}) \sim (0.7/\sqrt{2}, 0.7/\sqrt{2})$.

with the largest scale filamentary structure corresponding to the least dynamically evolved features of the field. Here the gas density was sampled over a cube of size 512^3. It was then smoothed over 8 pixels (300 kpc) and the persistent skeleton was computed from the logarithm of that smoothed field normalized to its standard deviation. Hence the persistence levels 0.06, 0.12, ..., 2 are in units of this root mean square.

Fig. C2 gives visual impression of the corresponding structure of the skeleton as a function of these persistence levels: the skeleton has a tree-like structure, for which each level of lower persistence contributes smaller branches. Hence the persistence level of 0.5 used in the main text corresponds to a description of the main filaments of the simulation.

APPENDIX D: THE EFFECT OF SMOOTHING

Fig. D1 shows the effect of the smoothing of the velocity field before computing the vorticity on the alignment between the vorticity and the direction of the filament. The amplitude of the excess of alignment varies slightly with the smoothing scale, but the main conclusion that an excess of alignment is detected remains unchanged.

APPENDIX E: TIDAL-VORTICITY LOCKUP

Fig. E1 displays the probability distribution of the cosine of the angle between the vorticity and the eigenvectors of the tidal field tensor, $\cos\gamma$. The vorticity tends to be perpendicular to the minor axis (e_3) of the tidal tensor which corresponds to the axis along which material is collapsing fastest. It is qualitatively in agreement with Fig. 2 and with Libeskind et al. (2013) which focus, respectively, on the eigenvectors of the Hessian of the density, and the eigenvectors of the shear tensor. For the latter, the description is kinematic, rather than dynamical for the tidal field.

APPENDIX F: FOF HALO CATALOGUE

As mentioned in the main text, FOF is prone to spuriously link neighbouring structures which could bias the alignment of the spin and the vorticity. An additional criterion is therefore required to produce a trustworthy catalogue of haloes. Following Bett et al.

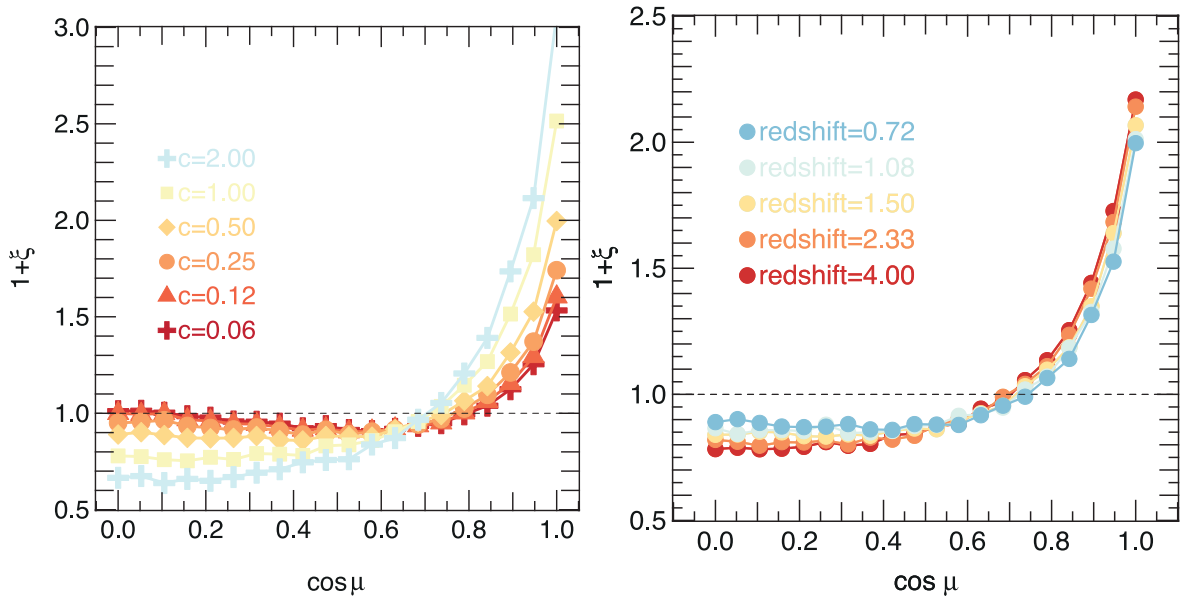


Figure C1. Left: the probability distribution of the cosine of the angle between the vorticity and the direction of the filament measured in $\mathcal{S}_{20}^{\rm cool}(z = 0.7)$ for different persistence threshold. The level of persistence of the main text corresponds to $c = 0.5$. Right: the probability distribution of the cosine of the angle between the vorticity and the direction of the skeleton, measured in $\mathcal{S}_{20}^{\rm cool}(0.7)$ for various redshifts as labelled. The amplitude of the correlation decreases with cosmic time.

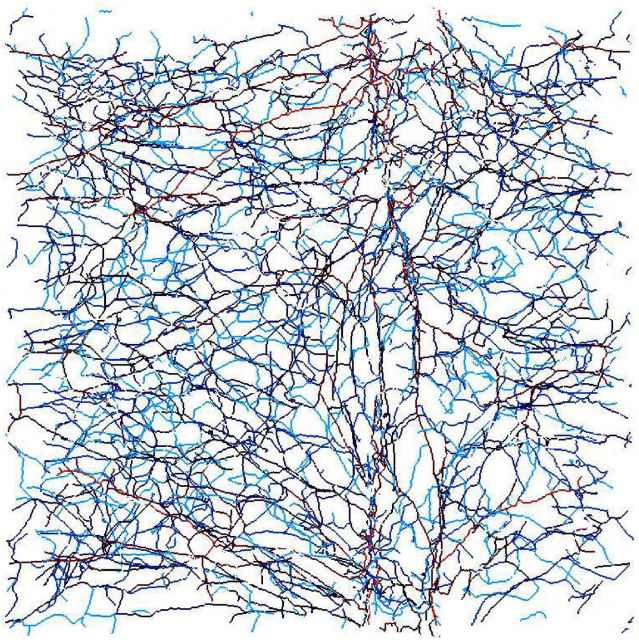


Figure C2. The skeleton measured in $\mathcal{S}_{20}^{\rm cool}(0.7)$ for increasing persistence threshold, 0.06, 0.12,. . . , 2, from light blue to red; the skeleton has tree-like structure where the main branches correspond to the most persistent ones. The level of persistence of the main text corresponds to the dark blue and red branches.

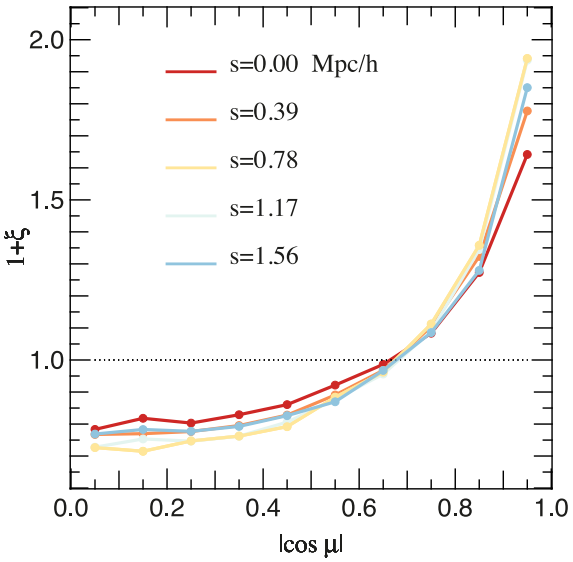


Figure D1. The probability distribution of the cosine of the angle between the vorticity and the direction of the filament, measured in $\mathcal{S}_{100}^{\rm CDM}$ for various smoothing scales of the velocity field before computation of the vorticity. Smoothing scales are expressed in h^{-1} Mpc. The smoothing scale adopted in the main text is 0.39 h^{-1} Mpc.

(2007), we proceed using the distribution of the spin parameter defined by Peebles (1969): $\boldsymbol{\lambda} = J|E|^{1/2}/GM_{\rm h}^{5/2}$, where J is the magnitude of the spin, E is the total energy of the halo, G is the gravitational constant and $M_{\rm h}$ is the halo mass.

Fig. F1 shows the average normalized histogram of the logarithm of the spin parameter for the haloes in the simulations set $\mathcal{S}_{50}^{\rm CDM}$. At high spin we clearly see a long tail, up to $\lambda = 238.2$, due to spurious linking of the structures. We use the analytical model proposed in Bett et al. (2007) to fit the $\log\lambda$-distribution: $P(\log\lambda) = A(\lambda/\lambda_0)^3\exp\left[-\alpha(\lambda/\lambda_0)^{3/\alpha}\right]$, where $A = 3\ln 10\,\alpha^{\alpha-1}/\Gamma(\alpha)$, with the values $\lambda_0 = 0.0341$ and $\alpha = 2.98$ which are providing the best fit. These values are in good agreement

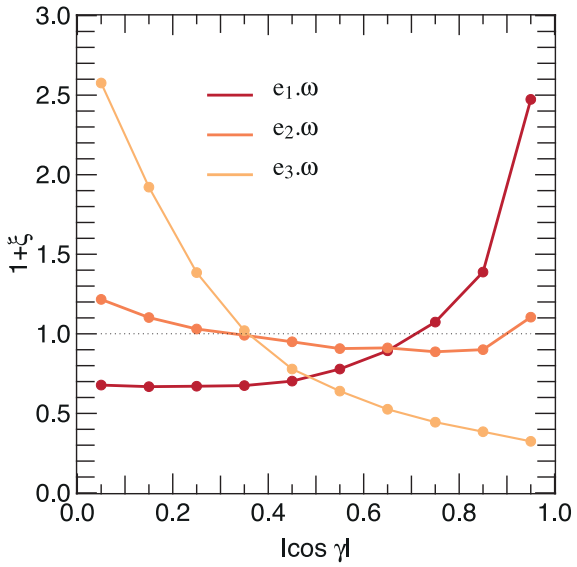


Figure E1. The probability distribution of the cosine of the angle between the vorticity and the eigenvectors of the tidal tensor, measured in $\mathcal{S}_{100}^{\rm CDM}$. The vorticity tends to be perpendicular to the minor axis (e_3) of the tidal tensor: the excess of probability to have $|\cos\theta|$ in [0, 0.5] (i.e. $60 \leq \theta \leq 90°$) is 50 per cent relative to random orientations. The vorticity tends also to be aligned with the major axis (e_1): the excess probability to have $|\cos\theta|$ in [0.5, 1] (i.e. $0 \leq \theta \leq 60°$) is 25 per cent relative to random orientations. e_3 corresponds to the axis along which material is collapsing fastest.

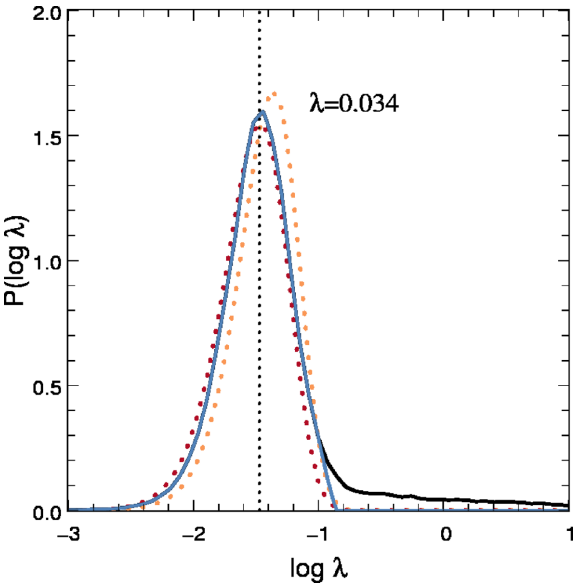


Figure F1. The normalized histogram of the logarithm of the spin parameter in the simulations set $\mathcal{S}_{50}^{\rm CDM}$. The median value of λ is 0.034 ± 0.0005 for the cleaned catalogue. The solid black line is the normalized distribution for all haloes. Notice a tail at high spin parameter which corresponds to spuriously linked structures. The solid blue line is the distribution for haloes with $\lambda \leq 0.12$, with the same normalization as for all haloes. The dotted lines are the analytical fit described in the text, red is our best fit to the distribution, and orange is the best fit found by Bett et al. (2007).

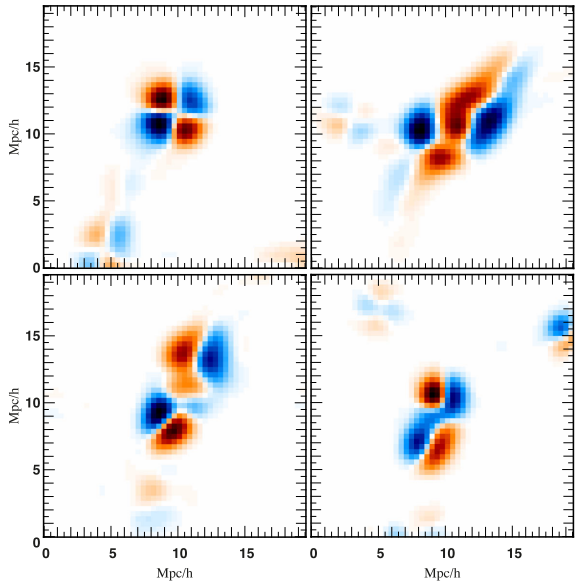


Figure G1. Different kinds of vorticity cross-sections.

with those found by Bett et al. (2007) ($\lambda_0 = 0.043$ and $\alpha = 2.51$), though their way to clean their catalogue (TREEall) is more sophisticated, in particular by taking into account an additional condition on energy. They showed also that the minimal number of particles per halo, $N_{\rm p}$, clearly affects the λ-distribution only for $N_{\rm p}$ lower than 100. Above this threshold, the change in the median value of λ stays lower than 10 per cent. Consequently, we keep in our catalogue only haloes with more than 100 particles. These haloes are then selected through a cut in λ. We find that removing haloes with $\lambda \geq 0.12$ best fits the adopted analytical model. Removed haloes represent 9.4 ± 1.2 per cent of the total population. Inspection of some of these removed haloes shows that they generally are multiple objects. We are left with around 5000 haloes in each 50 h^{-1} Mpc box of the $\mathcal{S}_{50}^{\rm CDM}$ simulations set.

We then quantify how the cut in λ affects the vorticity–spin alignment results. Considering three different catalogues with three different cuts in λ ($\lambda < 0.08$, $\lambda < 0.12$, $\lambda < 0.2$) we look for each catalogue and for each bin of mass at the quantity $(\zeta_{\rm tot} - \zeta_{\rm cut})/(1 + \zeta_{\rm tot})$ where $1 + \zeta_{\rm cut}$ is the excess of alignment in the reduced catalogue and $1 + \zeta_{\rm tot}$ in the full catalogue. This difference is always <5 per cent. We conclude that including or not the misidentified structures does not significantly change the measure of the spin alignment with the vorticity.

APPENDIX G: ZOOLOGY OF CAUSTICS

Fig. G1 shows a bundle of cross-sections of vorticity computed as in Fig. 3.

APPENDIX H: DEFINING ZERO VORTICITY

The algorithm DISPERSE introduced by Sousbie (2011) is used to defined the density walls and the contours of minimal vorticity. The density walls are computed as being the ascending two–manifolds of the skeleton calculated on the density field. The contours of minimal vorticity are defined as being the descending two–manifolds of the skeleton calculated on the norm of the vorticity

field. Since the vorticity is really well defined only on the neighbourhood of caustics, a mask is applied when the walls are computed, which covers all the regions of space where the density is lower than 10 per cent of the maximum density and the vorticity lower than 10 per cent of the maximum vorticity. The results of the computation of the density walls and minimal vorticity contours are tessellations, which means sets of triangles. For each triangle in the minimal vorticity tessellation we find its nearest neighbours in the density tessellation. Smoothing is achieved by averaging the position of each vertex with that of its direct neighbours. A smoothing coefficient $S = N$ means that this operation is repeated N times. The cosine between the normals of both triangles is then calculated.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **444,** 1453–1468 (2014) doi:10.1093/mnras/stu1227

Dancing in the dark: galactic properties trace spin swings along the cosmic web

Y. Dubois,[1,2,3,4★] C. Pichon,[1,2,4,5] C. Welker,[1,2,3] D. Le Borgne,[1,2] J. Devriendt,[3,6] C. Laigle,[1,2] S. Codis,[1,2] D. Pogosyan,[7] S. Arnouts,[8] K. Benabed,[1,2] E. Bertin,[1,2] J. Blaizot,[6] F. Bouchet,[1,2] J.-F. Cardoso,[1,2] S. Colombi,[1,2] V. de Lapparent,[1,2] V. Desjacques,[9] R. Gavazzi,[1,2] S. Kassin,[10] T. Kimm,[11] H. McCracken,[1,2] B. Milliard,[8] S. Peirani,[1,2] S. Prunet,[1,2,12] S. Rouberol,[1,2] J. Silk,[1,2,3,13] A. Slyz,[3] T. Sousbie,[1,2] R. Teyssier,[14] L. Tresse,[8] M. Treyer,[8] D. Vibert[8] and M. Volonteri[1,2,3,15]

[1] *Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75005 Paris, France*
[2] *CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, F-75014 Paris, France*
[3] *Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*
[4] *KITP Kohn Hall, 4030 University of California Santa Barbara, CA 93106-4030, USA*
[5] *Institute of Astronomy and Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK*
[6] *Observatoire de Lyon, UMR 5574, 9 avenue Charles André, F-69561 Saint Genis Laval, France*
[7] *Department of Physics, University of Alberta, 11322-89 Avenue, Edmonton, Alberta T6G 2G7, Canada*
[8] *Aix Marseille Université, CNRS, Laboratoire d'Astrophysique de Marseille, UMR 7326, 38 rue F. Joliot-Curie, F-13388 Marseille, France*
[9] *Département de Physique Théorique, Université de Genève, 24 quai Ernest Ansermet, CH-1211 Genève, Switzerland*
[10] *Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[11] *Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[12] *Canada–France–Hawaii Telescope Corporation, 65-1238 Mamalahoa Hwy, Kamuela, HI 96743, USA*
[13] *Department of Physics and Astronomy, The Johns Hopkins University Homewood Campus, Baltimore, MD 21218, USA*
[14] *Institute für Theoretische Physik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[15] *Astronomy Department, University of Michigan, Ann Arbor, MI 48109, USA*

Accepted 2014 June 19. Received 2014 June 19; in original form 2014 January 29

ABSTRACT

A large-scale hydrodynamical cosmological simulation, Horizon-AGN, is used to investigate the alignment between the spin of galaxies and the cosmic filaments above redshift 1.2. The analysis of more than 150 000 galaxies per time step in the redshift range $1.2 < z < 1.8$ with morphological diversity shows that the spin of low-mass blue galaxies is preferentially aligned with their neighbouring filaments, while high-mass red galaxies tend to have a perpendicular spin. The reorientation of the spin of massive galaxies is provided by galaxy mergers, which are significant in their mass build-up. We find that the stellar mass transition from alignment to misalignment happens around 3×10^{10} $M_{\odot}$. Galaxies form in the vorticity-rich neighbourhood of filaments, and migrate towards the nodes of the cosmic web as they convert their orbital angular momentum into spin. The signature of this process can be traced to the properties of galaxies, as measured relative to the cosmic web. We argue that a strong source of feedback such as active galactic nuclei is mandatory to quench *in situ* star formation in massive galaxies and promote various morphologies. It allows mergers to play their key role by reducing post-merger gas inflows and, therefore, keeping spins misaligned with cosmic filaments.

Key words: methods: numerical – galaxies: evolution – galaxies: formation – galaxies: kinematics and dynamics – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

Theoretical models of structure formation by gravitational instability and numerical simulations have predicted that small fluctuations from the early Universe lead to the formation of a large-scale

★ E-mail: dubois@iap.fr

cosmic web made of clustered haloes, filaments, sheets and voids (e.g. Zeldovich, Einasto & Shandarin 1982; Klypin & Shandarin 1983; Blumenthal et al. 1984; Davis et al. 1985). The resulting properties of the Universe's large-scale structure are the interplay of the planar local collapse, as emphasized in Zel'dovich (1970) (see also Shandarin & Zeldovich 1989) and the inherent structure of the Gaussian initial density and velocity shear fields, leading to the cosmic web picture of dense peaks connected by filaments, framing the honeycomb-like structure of walls (Bardeen et al. 1986; Bond, Kofman & Pogosyan 1996).

The extension of the Center for Astrophysics redshift survey (Huchra et al. 1983) gave spectacular observational evidence (de Lapparent, Geller & Huchra 1986; Geller & Huchra 1989) for this picture, triggering a renewed interest for such large-scale galaxy surveys (Colless et al. 2001; Tegmark et al. 2004).

Modern simulations have established a tight connection between the geometry and dynamics of the large-scale structure of matter, on the one hand, and the evolution of the physical properties of forming galaxies, on the other. Observational information on the morphology of galaxies and its dependence on environment is routinely becoming available for galaxies up to redshift 2 and beyond (Abraham et al. 2007; Oesch et al. 2010; Lee et al. 2013). Matched samples at low and high redshifts allow for the study of the evolution of many physical properties of galaxies for most of the history of our Universe in unprecedented detail. A key question formulated decades ago is nevertheless not satisfactorily answered: what properties of galaxies are driven by the cosmic environment?

There is ample literature (e.g. Hoyle 1949; Peebles 1969; Doroshkevich 1970; White 1984; Schaefer 2009) on the tidal torque theory (TTT). It aims to explain the early acquisition of the spin[1] of haloes, in the regime where the dynamics is well described by the Zel'dovich approximation, and when it is legitimate to assume that the tidal and the inertia tensors are uncorrelated. Within this framework, TTT predicts that the spin of haloes should be perpendicular to the direction of filaments.

Using *N*-body simulations, which model only dark matter (DM), Hahn et al. (2007a) and Zhang et al. (2009) found halo spins preferentially oriented perpendicular to the filaments independent of halo mass. Hatton & Ninin (2001) claimed to detect an alignment between spin and filament, while Faltenbacher et al. (2002) measured a random orientation of the spins of haloes in the plane perpendicular to the filaments. More recently, a consensus emerged when several works (Aubert, Pichon & Colombi 2004; Bailin & Steinmetz 2005; Aragón-Calvo et al. 2007; Hahn et al. 2007b; Paz, Stasyszyn & Padilla 2008; Sousbie et al. 2008; Libeskind et al. 2012; Trowland, Lewis & Bland-Hawthorn 2013) reported that large-scale structures – filaments and sheets – influence the direction of the angular momentum of haloes in a way originally predicted by Sugerman, Summers & Kamionkowski (2000) and Lee & Pen (2000). These studies pointed towards a mass-dependent orientation of the spin, arguing for the first time that the spin of high-mass haloes tends to lie perpendicular to their host filament, whereas low-mass haloes have a spin preferentially aligned with it. Nevertheless, the detected correlation remained weak and noisy until Codis et al. (2012) confirmed it. They quantified a redshift-dependent mass transition $M_{\rm tr,\,h}$, separating aligned from perpendicular haloes, and interpreted the origin of the transition in terms of large-scale cosmic flows. Codis et al. (2012) found that high-mass haloes have their spins perpendicular to the filament because they are the results of mergers, a scenario suggested earlier by Aubert et al. (2004, see also Bailin & Steinmetz 2005). Low-mass haloes are not the products of mergers and acquire their mass by gas accretion in the vorticity-rich neighbourhood of filaments, which explain why their spins are initially parallel to the filaments (Laigle et al. 2013; Libeskind et al. 2013a).

Tempel, Stoica & Saar (2013) recently found tentative evidence of such alignments in the Sloan Digital Sky Survey (SDSS) with an orthogonality for elliptical galaxies and a weak alignment for spiral galaxies (see also Tempel & Libeskind 2013). Zhang et al. (2013) found that the major axis of red galaxies is parallel to their host filaments and is the same for blue galaxies albeit with a weaker signature. Similar measurements have been done for galaxies and walls; there is evidence that the spin of galaxies also lies within the walls in which they are contained (Trujillo, Carretero & Patiri 2006).

Besides those attempts to relate the spins of galaxies with the cosmic structure, much observational effort has been made to control the level of intrinsic alignments of galaxies as a potential source of systematic errors in weak gravitational lensing measurements (e.g. Croft & Metzler 2000; Heavens, Refregier & Heymans 2000; Hirata & Seljak 2004). Such alignments are believed to be the major source of systematics of the future generation of lensing surveys like *Euclid* or Large Synoptic Survey Telescope (LSST). Direct measurements of the alignment of the projected light distribution of galaxies in wide-field imaging data seem to agree on a contamination at a level between a few per cent and $\sim$10 per cent of the shear correlation functions, although the amplitude of the effect depends on the population of galaxies considered (e.g. Lee & Pen 2002; Joachimi et al. 2013). Given this dependence, it is difficult to use DM-only simulations as the sole resource to predict and control intrinsic alignments despite some success with the addition of a semi-analytical model prescription (e.g. Joachimi et al. 2013). The inherently anisotropic nature of the large-scale structure and its complex imprint on the shapes and spins of galaxies may prevent isotropic approaches from making accurate predictions.

Very few attempts have been made to probe the degree of correlation between *galaxy* spins and their embedding cosmic web using hydrodynamical cosmological simulations. Hahn, Teyssier & Carollo (2010) simulated the vicinity of a large-scale cosmic filament and found that – at odds with the results presented in our paper – the spin of galaxies within high-mass haloes is aligned with the filament while the spin of galaxies in low-mass haloes is perpendicular to the filament. Gay et al. (2010) focused on the colour gradients relative to the cosmic web using the Horizon-MareNostrum simulation (Devriendt et al. 2010) which did not display much morphological diversity. They found evidence of metallicity gradients towards and along the filaments and nodes of the cosmic web. Danovich et al. (2012) also studied the feeding of massive galaxies at high redshift through cosmic streams using the Horizon-MareNostrum simulation. They found that galaxies are fed by one dominant stream (with a tendency to be fed by three major streams), streams tend to be co-planar (in the stream plane), and that there is a weak correlation between spin of the galaxy and spin of the stream plane at the virial radius, which suggests an angular momentum exchange at the interface between streams and galaxies (see also Tillson et al. 2012).

In this paper, our focus is on the influence of the cosmic web as an anisotropic vector of the gas mass and angular momentum which ultimately shape galaxies. Our purpose is to determine if the mass-dependent halo spin–filament correlations of Codis et al. (2012) can be recovered via the morphology and physical

[1] Hereafter, the spin is the angular momentum unit vector for simplicity.

properties of *simulated galaxies*. We aim to test these findings on a state-of-the-art hydrodynamical simulation, the so-called Horizon-AGN simulation, which produced over 150 000 resolved galaxies displaying morphological diversity by redshift $z = 1.2$, in order to identify the effect of the environmentally driven spin acquisition on morphology and to probe the tendency of galaxies to align or misalign with the cosmic filaments as a function of galactic properties.

The paper is organized as follows. Section 2 describes the numerical set-up of our simulation, the post-processing of galaxy properties and filament tracing. Section 3 presents the excess probability of alignment as a function of morphological tracers and investigates its redshift and spatial evolution. Section 4 discusses the origin of the observed (mis)alignment between galactic spins and filaments. We finally conclude in Section 5.

2 THE VIRTUAL DATA SET

In this section, we describe the Horizon-AGN simulation (Section 2.1), how galaxies are identified in it and how the virtual observables are derived (Section 2.2). We conclude this section with a description of the tools used to compare the spins of galaxies to the orientation of the cosmic web (Section 2.3).

2.1 The Horizon-AGN simulation

2.1.1 The code and initial conditions

We adopt a standard Λ cold dark matter cosmology with total matter density $\Omega_{\rm m} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, amplitude of the matter power spectrum $\sigma_8 = 0.81$, baryon density $\Omega_{\rm b} = 0.045$, Hubble constant $H_0 = 70.4\,{\rm km\,s^{-1}\,Mpc^{-1}}$ and $n_{\rm s} = 0.967$ compatible with the *Wilkinson Microwave Anisotropy Probe* 7 cosmology (Komatsu et al. 2011). The values of this set of cosmological parameters are compatible with those of the recent *Planck* results within a 10 per cent relative variation (Planck Collaboration 2013). The size of the box is $L_{\rm box} = 100\,h^{-1}$ Mpc with 1024^3 DM particles, which results in a DM mass resolution of $M_{\rm DM,\,res} = 8 \times 10^7\,{\rm M}_\odot$. The initial conditions have been produced with the MPGRAFIC software (Prunet et al. 2008). The simulation was run down to $z = 1.2$ and used 4 million CPU hours.

The Horizon-AGN simulation is run with the adaptive mesh refinement code RAMSES (Teyssier 2002). The evolution of the gas is followed using a second-order unsplit Godunov scheme for the Euler equations. The HLLC Riemann solver (Toro, Spruce & Speares 1994) with MinMod total variation diminishing scheme is used to reconstruct the interpolated variables from their cell-centred values. Collisionless particles (DM and star particles) are evolved using a particle-mesh solver with a cloud-in-cell interpolation. The initial mesh is refined up to $\Delta x = 1$ kpc (seven levels of refinement). This is done according to a quasi-Lagrangian criterion: if the number of DM particles in a cell is more than 8, or if the total baryonic mass in a cell is eight times the initial DM mass resolution, a new refinement level is triggered. In order to keep the minimum cell size approximately constant in physical units, we allow a new maximum level of refinement every time the expansion scale factor doubles (i.e. at $a_{\rm exp} = 0.1$, 0.2, 0.4 and 0.8).

2.1.2 Gas cooling and heating

Gas is allowed to cool by H and He cooling with a contribution from metals using a Sutherland & Dopita (1993) model down to 10^4 K. Heating from a uniform UV background takes place after redshift $z_{\rm reion} = 10$ following Haardt & Madau (1996). Metallicity is modelled as a passive variable for the gas, and its amount is modified by the injection of gas ejecta during supernova (SN) explosions and stellar winds. We also account for the release of various chemical elements synthesized in stars and released by stellar winds and SNe: O, Fe, C, N, Mg and Si. However, they do not contribute separately to the cooling curve (the ratio between each element is taken to be solar for simplicity) but can be used to probe the distribution of the various metal elements. The gas follows an equation of state for an ideal monoatomic gas with an adiabatic index of $\gamma = 5/3$.

2.1.3 Star formation and stellar feedback

The star formation process is modelled with a Schmidt law: $\dot\rho_* = \epsilon_*\rho/t_{\rm ff}$, where $\dot\rho_*$ is the star formation rate (SFR) density, $\epsilon_* = 0.02$ (Kennicutt 1998; Krumholz & Tan 2007) the constant star formation efficiency and $t_{\rm ff}$ the local free-fall time of the gas. Star formation is allowed in regions which exceed a gas hydrogen number density threshold of $n_0 = 0.1\,{\rm H\,cm^{-3}}$ following a Poissonian random process (Rasera & Teyssier 2006; Dubois & Teyssier 2008) with a stellar mass resolution of $M_* = \rho_0\Delta x^3 \simeq 2 \times 10^6\,{\rm M}_\odot$. The gas pressure is artificially enhanced above $\rho > \rho_0$ assuming a polytropic equation of state $T = T_0(\rho/\rho_0)^{\kappa-1}$ with polytropic index $\kappa = 4/3$ to avoid excessive gas fragmentation and mimic the effect of stellar heating on the mean temperature of the interstellar medium (Springel & Hernquist 2003). Feedback from stars is explicitly taken into account assuming a Salpeter (1955) initial mass function (IMF) with a low-mass (high-mass) cut-off of $0.1\,{\rm M}_\odot$ ($100\,{\rm M}_\odot$), as described in detail in Kimm et al. (in preparation). Specifically, the mechanical energy from Type II SNe and stellar winds is taken from STARBURST99 (Leitherer et al. 1999, 2010), and the frequency of Type Ia SN explosions is computed following Greggio & Renzini (1983).

2.1.4 Feedback from black holes

The same 'canonical' active galactic nucleus (AGN) feedback modelling employed in Dubois et al. (2012a) is used here. Black holes (BHs) are created where the gas mass density is larger than $\rho > \rho_0$ with an initial seed mass of $10^5\,{\rm M}_\odot$. In order to avoid the formation of multiple BHs in the same galaxy, BHs are not allowed to form at distances less than 50 kpc from each other. The accretion rate on to BHs follows the Bondi–Hoyle–Lyttleton rate $\dot M_{\rm BH} = 4\pi\alpha G^2 M_{\rm BH}^2\bar\rho/(\bar c_{\rm s}^2 + \bar u^2)^{3/2}$, where $M_{\rm BH}$ is the BH mass, $\bar\rho$ is the average gas density, $\bar c_{\rm s}$ is the average sound speed, $\bar u$ is the average gas velocity relative to the BH velocity and α is a dimensionless boost factor with $\alpha = (\rho/\rho_0)^2$ when $\rho > \rho_0$ and $\alpha = 1$ otherwise (Booth & Schaye 2009) in order to account for our inability to capture the colder and higher density regions of the interstellar medium. The effective accretion rate on to BHs is capped at the Eddington accretion rate: $\dot M_{\rm Edd} = 4\pi G M_{\rm BH} m_{\rm p}/(\epsilon_{\rm r}\sigma_{\rm T}c)$, where $\sigma_{\rm T}$ is the Thompson cross-section, c is the speed of light, $m_{\rm p}$ is the proton mass and $\epsilon_{\rm r}$ is the radiative efficiency, assumed to be equal to $\epsilon_{\rm r} = 0.1$ for the Shakura & Sunyaev (1973) accretion on to a Schwarzschild BH.

The AGN feedback is a combination of two different modes, the so-called *radio* mode operating when $\chi = \dot M_{\rm BH}/\dot M_{\rm Edd} < 0.01$ and the *quasar* mode active otherwise. The quasar mode consists of an isotropic injection of thermal energy into the gas within a sphere of radius Δx, and at an energy deposition rate: $\dot E_{\rm AGN} = \epsilon_{\rm f}\epsilon_{\rm r}\dot M_{\rm BH}c^2$. In this equation, $\epsilon_{\rm f} = 0.15$ is a free parameter chosen to reproduce the scaling relations between BH mass and galaxy

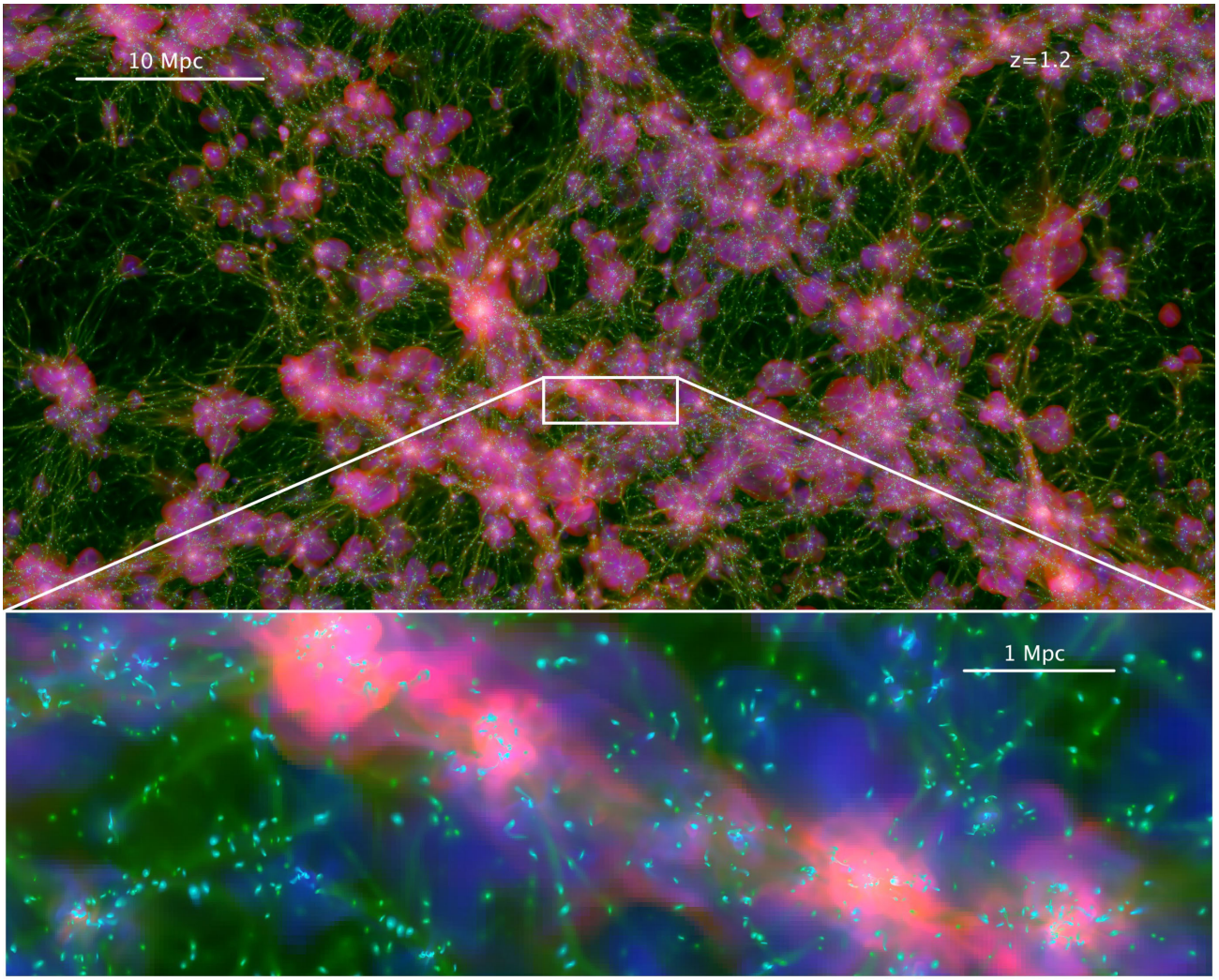


Figure 1. Projected maps of the Horizon-AGN simulation at $z = 1.2$ are shown. Gas density (green), gas temperature (red) and gas metallicity (blue) are depicted. The top image is $100\,h^{-1}$ Mpc across in comoving distance and covers the whole horizontal extent of the simulation and $25\,h^{-1}$ Mpc comoving in depth. The bottom image is a subregion where we see thin cosmic filaments as well as thicker filaments several Mpc long bridging shock-heated massive haloes and surrounded by a metal-enriched intergalactic medium. Physical scales are indicated on the figures in proper units.

properties (mass, velocity dispersion) and BH density in our local Universe (see Dubois et al. 2012a). At low accretion rates, the radio mode deposits AGN feedback energy into a bipolar outflow with a jet velocity of 10^4 km s^{-1}. The outflow is modelled as a cylinder with a cross-sectional radius Δx and height $2\,\Delta x$ following Omma et al. (2004) (more details are given in Dubois et al. 2010). The efficiency of the radio mode is larger than the quasar mode with $\epsilon_{\rm f} = 1$.

A projected map of half the simulation volume and a smaller subregion is shown in Fig. 1. Gas density, gas temperature and gas metallicity are depicted. One can discern the large-scale pattern of the cosmic web, with filaments and walls surrounding voids and connecting haloes. Massive haloes are filled with hot gas, and feedback from SNe and AGN pours warm and metal-rich gas in the diffuse intergalactic medium. As demonstrated in Dubois et al. (2013), the modelling of AGN feedback is critical to create early-type galaxies and provide the sought morphological diversity (see Fig. 2 for a snippet of the galaxy sample of the simulation) in hydrodynamical cosmological simulations (see e.g. Croton et al. 2006 for semi-analytical models).

2.2 Mock observations of galaxies

We describe how we produce various observables that can be compared qualitatively with data from modern observational surveys. In this paper, we focus on observables which are known to correlate with the Hubble type of galaxies, namely mass, V/σ, colour, morphological parameters like Gini and M_{20}, and age.

2.2.1 Identifying and segmenting galaxies

Galaxies are identified with the AdaptaHOP finder (Aubert et al. 2004, updated to its recent version by Tweed et al. 2009 for building merger trees) which directly operates on the distribution of star particles. A total of 20 neighbours are used to compute the local density of each particle, a local threshold of $\rho_{\rm t} = 178$ times the average total matter density is applied to select relevant densities, and the force softening (minimum size below which substructures are considered irrelevant) is $\sim$2 kpc. Only galactic structures identified with more than 50 particles are considered. It allows for a clear separation of galaxies (defined as sets of star particles segmented

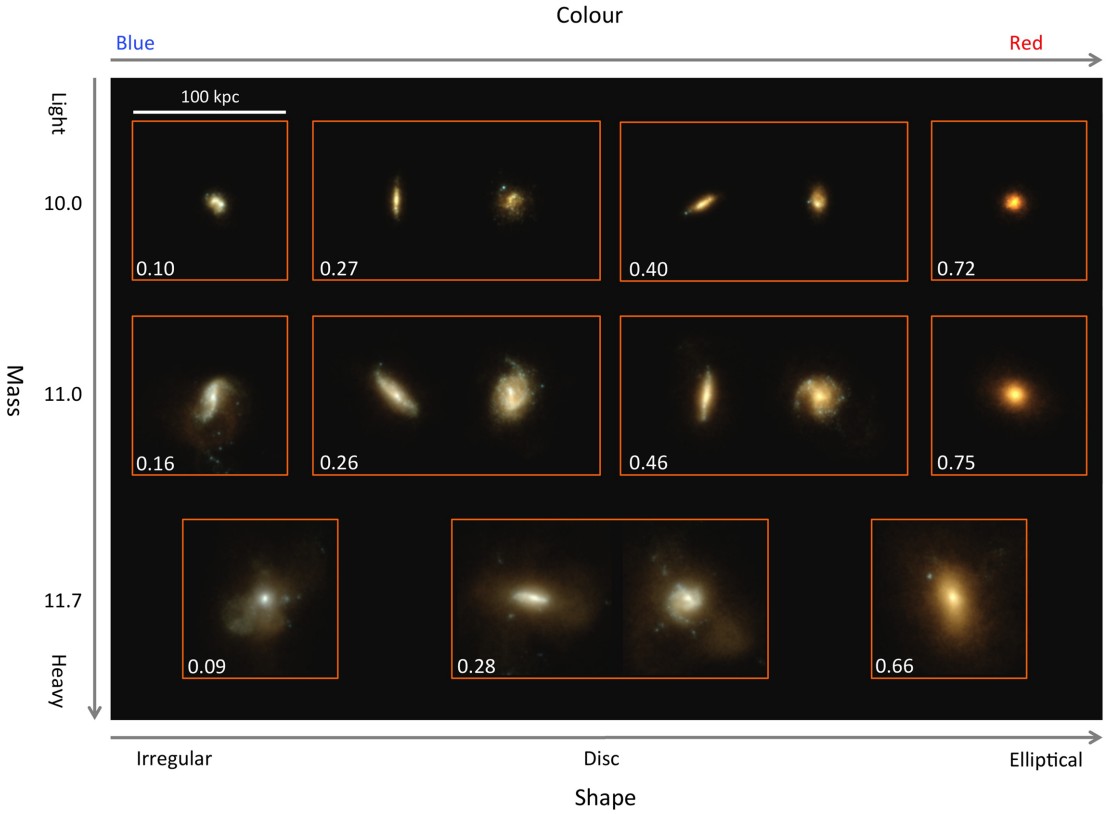


Figure 2. Stellar emission of a sample of galaxies in the Horizon-AGN simulation at $z = 1.3$ observed through rest-frame u, g and i filters. Extinction by dust is not taken into account. Each vignette size is 100 kpc vertically. The numbers on the left of the figure indicate the galaxy stellar mass in log solar mass units. The number in the bottom left of each vignette is the $g - r$ rest-frame colour, not corrected for dust extinction. Disc galaxies (galaxies in the centre of the figure) are shown edge-on and face-on.

by AdaptaHOP), including those in the process of merging. Catalogues of around ∼150 000 galaxies are produced for each redshift analysed in this paper from $z = 3$ to 1.2.

2.2.2 *Synthetic colours*

We compute the absolute AB magnitudes and rest-frame colours of galaxies using single stellar population models from Bruzual & Charlot (2003) assuming a Salpeter IMF. Each star particle contributes to a flux per frequency that depends on its mass, age and metallicity. The sum of the contribution from all stars is passed through the u, g, r and i filters from the SDSS. Fluxes are expressed as rest-frame quantities (i.e. that do not take into account the red-shifting of spectra). We also neglect the contribution to the reddening of spectra from internal (interstellar medium) or external (intergalactic medium) dust extinction. Once the flux in each waveband is obtained for a star particle, we build two-dimensional projected maps from single galaxies (satellites are excised with the galaxy finder), and we can sum up the total contribution of their stars to the total luminosity. A small sample of galaxies representative of the morphological variety in the simulation is shown in Fig. 2.

2.2.3 *Projected stellar kinematics*

For each galaxy, we build a field of view centred on the galaxy, which is made of 256 × 256 pixels over 100 kpc size (corresponding to a pixel size of 0.4 kpc or 0.05 arcsec at $z = 1.83$). We compute the luminosity-weighted velocity along the line of sight (arbitrary defined as the x-axis of the simulation):

$$\bar{v}_{\rm pixel} = \frac{\Sigma_i v_{{\rm los},i} I_{i,{\rm filter}}}{\Sigma_i I_{i,{\rm filter}}}, \quad (1)$$

where $v_{{\rm los},i}$ is the velocity along the line of sight of the ith star in the pixel considered and $I_{i,{\rm filter}}$ is the intensity in the corresponding filter bandwidth (u, g, r, i) of the ith star in the pixel considered. Then, the velocity dispersion along the line of sight is

$$\bar{\sigma}^2_{\rm pixel} = \frac{\Sigma_i v^2_{{\rm los},i} I_{i,{\rm filter}}}{\Sigma_i I_{i,{\rm filter}}} - \bar{v}^2_{\rm pixel}. \quad (2)$$

The velocity maps are then smoothed with a Gaussian kernel of 15 pixels. The position of the fastest (respectively slowest) pixel, which defines V for that galaxy, is then identified automatically and a 0.75 arcsec 'slit' is put across so as to interpolate through the kinematic major axis of the galaxy. The smoothed velocity dispersion map is also interpolated along the same axis, and the maximum of that curve defines σ (see Fig. 3 for example of a slow and a fast rotator). V/σ is then straightforwardly the corresponding ratio.

2.2.4 *Specific star formation rate*

The calculation of the SFR is done on stars as identified by the galaxy finder that belong to a given galaxy. To compute the SFR, we compute the amount of stars formed over the last 100 Myr. The choice of 100 Myr corresponds to a minimum measurable SFR

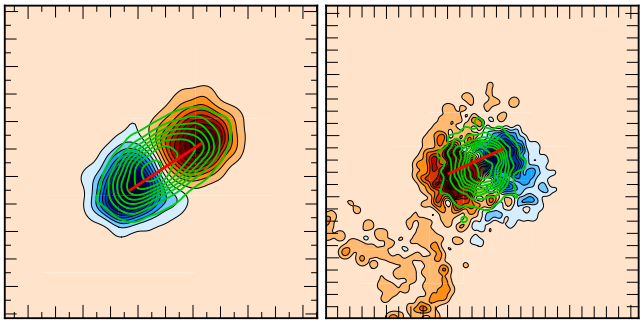


Figure 3. Example of integral field spectroscopy for the velocity of a fast rotator $V/\sigma > 1$ (left-hand panel) and a slow rotator $V/\sigma < 1$ (right-hand panel). The thick red line corresponds to the position of the slit, which is placed along the kinetic major axis. The green isocontours correspond to the velocity dispersion map. The size of the images is 50 kpc and the mean velocity (respectively dispersion) amplitude is 150 (respectively 75) km s^{-1} (left-hand panel) and 50 (respectively 100) km s^{-1} (right-hand panel).

of $M_*/100\,\mathrm{Myr} = 0.02\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$. The specific star formation rate (sSFR) is then calculated by diving the SFR by the galaxy stellar mass, M_s.

2.2.5 *Gini and M_{20}*

The morphology of each galaxy is often measured by two non-parametric parameters: the Gini (G; Abraham, van den Bergh & Nair 2003) and M_{20} (Lotz, Primack & Madau 2004). The Gini parameter is a non-parametric measure of the inequality of fluxes in pixels, ranging from zero (for a perfectly uniform image) to unity (for an image with all the flux in one pixel, for instance). M_{20} is the second-order momentum of light of the 20 per cent brightest pixels of a galaxy. It traces the spatial distribution of any bright nuclei, bars, spiral arms and off-centre star clusters. As shown in Lotz et al. (2004), galaxies with $M_{20} > -1.1$ mainly are extended objects with double or multiple nuclei, whereas low values of $M_{20}(< -1.6)$ are relatively smooth with a bright nucleus. Both parameters are known to correlate well with the concentration parameter (Abraham et al. 1994) for regular shapes, but they are better suited for disturbed morphologies because of their non-parametric nature. These two parameters have been used to characterize observations in the local universe and at high redshift (e.g. Lotz et al. 2004; Abraham et al. 2007; Wang et al. 2012; Lee et al. 2013), and they are well suited to analyse large samples of galaxies of mixed morphologies. In the local universe, galaxies with a high Gini value and a low M_{20} value are mainly ellipticals, whereas late-type galaxies and irregular have lower G and larger M_{20} values. Mergers tend to have large G and large M_{20} values.

Images in the i band are obtained from a segmentation of 3D objects with the galaxy finder. The images are rebinned to 64 × 64 pixels for a 100 kpc size image in order to avoid star particles appearing as individual pixels. Then, as in Lee et al. (2013), we measure the Petrosian radius with an elliptical aperture which is obtained as in SExtractor (Bertin & Arnouts 1996) from the second-order moment of light. The Petrosian semi-major axis a_p is such that the ratio of the surface brightness at a_p over the mean surface brightness within a_p is decreasing at a_p and becomes smaller than 0.2. In practice, we fit a spline to the surface brightness ratio profile and find the zero of the function $\mu(a_\mathrm{p})/\mu(< a_\mathrm{p}) - 0.2$. Galaxies with a_p smaller than 2 pixels are filtered out: they are almost always associated with low-mass galaxies ($M_\mathrm{s} < 10^{9.2}\,\mathrm{M}_\odot$) with few star particles, and the G and M_{20} parameters are very uncertain for these objects. We also filter out galaxies less massive than $M_\mathrm{s} < 10^{9.5}\,\mathrm{M}_\odot$ for which we suffer the most from resolution effects. A description of the bivariate distributions of G, M_{20} and stellar mass is given in Appendix A.

2.2.6 *Ages*

The mean ages of galaxies are obtained through the summation of the mass-weighted age of star particles belonging to the galaxy.

2.2.7 *Spin of galaxies*

To compute the spin of galaxies, we compute the total angular momentum of their stars with respect to the particle of maximum density (centre of the galaxy) from the smoothed stellar density constructed with the AdaptaHOP algorithm.

We have also tested the effect of grid-locking on the Cartesian axes of the box (a common issue of Cartesian-based Poisson solvers for which a numerical anisotropy in the force calculation arises; see e.g. Hockney & Eastwood 1981) in Appendix B for galaxies and filaments.

2.3 Tracing large-scale structures via the skeleton

In order to quantify the orientation of galaxies relative to the cosmic web, we use a geometric three-dimensional ridge extractor well suited to identify filaments, called the 'skeleton'. A gas density cube of 512^3 pixels is drawn from the simulation and Gaussian-smoothed with a length of $3\,h^{-1}$ Mpc comoving chosen so as to trace large-scale filamentary features. Two implementations of the skeleton, based on 'watershed' (Sousbie, Colombi & Pichon 2009) and 'persistence' (Sousbie, Pichon & Kawahara 2011), were implemented, without significant difference for the purpose of this investigation. The first method identifies ridges as the boundaries of walls which are themselves the boundaries of voids. The second one identifies ridges as the 'special' lines connecting topologically robust (filament-like) saddle points to peaks.

Fig. 4 shows a slice of $25\,h^{-1}$ Mpc of the skeleton colour coded by logarithmic density, along with galaxies contained within that slice. The clustering of the galaxies follows quite closely the skeleton of the gas, i.e. the cosmic filaments. Note that, on large scales, the skeleton built from the gas is equivalent to that built from the DM as the gas and DM trace each other closely. The rest of the paper is devoted to studying the orientation of the spin of these galaxies relative to the direction of the nearest skeleton segment. In practice, an octree is built from the position of the mid-segment of the skeleton to speed up the association of the galaxy position to its nearest skeleton segment. It was checked that our results were not sensitive to how many such segments were considered to define the local direction of the skeleton. The orientation of the segment of the skeleton is used to define the relative angle between the filament and the spin of the galaxy. The segments are also tagged with their curvilinear distance to the closest node (where different filaments merge), which allows us to study the evolution of this (mis)alignment *along* the cosmic web. Appendix B investigates the effect of grid-locking of the skeleton's segments in the Horizon-AGN simulation. Large-scale filaments, defined from the skeleton, do not show any alignment with the grid.

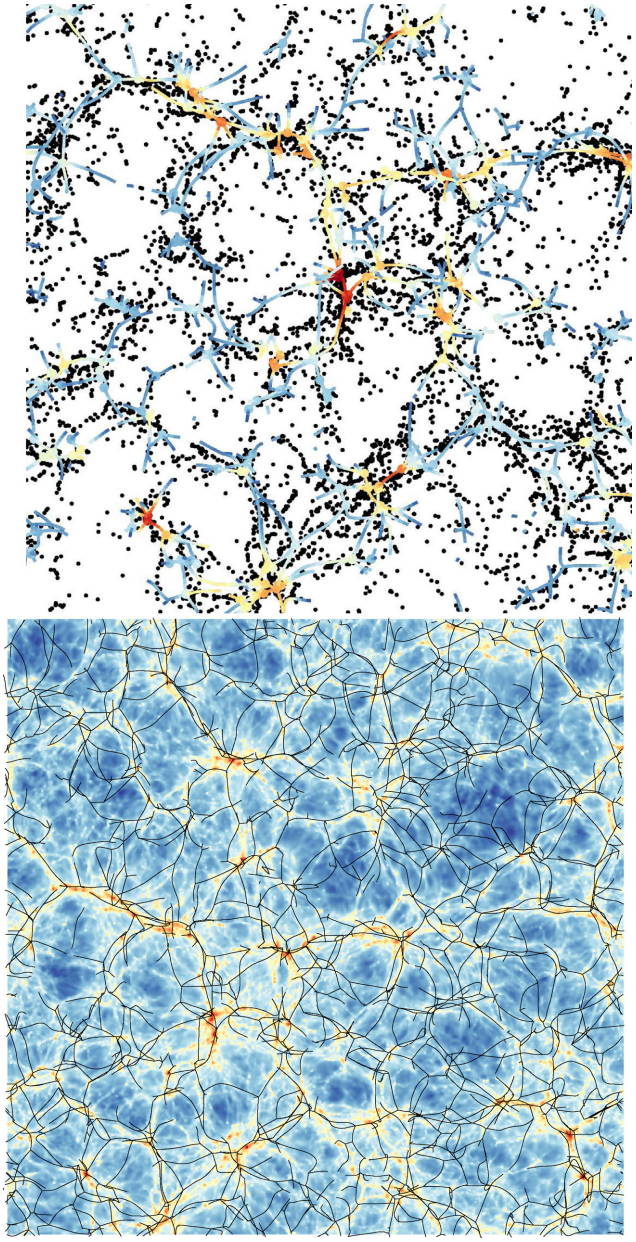


Figure 4. Top: projection along the z-axis of the Horizon-AGN gas skeleton (colour coded by logarithmic density as red–yellow–blue–white from high density to low density) at redshift $z = 1.83$ of a slice of $25\,h^{-1}$ Mpc on the side and $10\,h^{-1}$ Mpc thickness. Galaxies are superimposed as black dots. The clustering of the galaxies follows the skeleton quite closely. Bottom: larger view of the skeleton on top of the projected gas density. This paper quantifies orientation of the galaxies relative to the local anisotropy set by the skeleton.

3 SPIN SWING EVOLUTION

This paper focuses on the orientation of the spin of galaxies relative to the filaments in which they are embedded and the cosmic evolution. Specifically, we aim to see if its evolution can be traced via physical and morphological tracers. We investigate in Section 3.1 how this orientation varies with different tracers of the Hubble type of galaxies, namely stellar mass, V/σ, sSFR, colour, metallicity, age, M_{20} and Gini on our sample of 150 000 galaxies. Afterwards, in Section 3.2, we quantify how this alignment varies as a function of distance to the filaments and along the filaments to the nodes of the cosmic web. We study the cosmic evolution of the alignment of the spin of galaxies and filaments in Section 3.3.

3.1 Alignment of galaxies and filaments

We measure the statistical signature of the (mis)alignment of galaxies with their closest filament segment. The alignment is defined as the angle θ between the spin of the stellar component and the direction of the filamentary segment. Recall that the spin of the galaxy is obtained by removing merging substructures using the galaxy finder, and computing the net angular momentum of its stars with respect to its centre (defined as the point of highest stellar density). Note that the filament segments are assumed to have no polarity. Hence, we impose the angle θ has a $\pi/2$ symmetry, and is expressed in terms of $\cos\theta = [0, 1]$.

Fig. 5 shows the resulting probability density function (PDF), $1 + \xi$, at $z = 1.83$, where ξ is the excess probability of $\cos\theta$ in bins of various quantities: mass, kinematics, sSFR, colour, metallicity, age, M_{20} and Gini. A uniform PDF (i.e. random orientations of galaxies relative to their filament) is represented as a dashed line for comparison. Galaxies with mass below $M_{\rm s} < 10^9\,{\rm M}_\odot$ are removed from the calculation, except for investigation of alignment as a function of mass.

More massive galaxies tend to have their spin preferentially perpendicular to their filament, while less massive ones have their spin preferentially parallel. A transition occurs around a stellar mass of $M_{\rm tr,s} = 3 \times 10^{10}\,{\rm M}_\odot$. This value is fully consistent with earlier findings of a mass transition for the orientation of the spin of haloes of $M_{\rm tr,h} = 5 \times 10^{11}\,{\rm M}_\odot$ at that redshift (Codis et al. 2012) and suggested by the galaxy–halo mass relation determined by abundance matching techniques (Moster, Naab & White 2013). Using the full redshift sample, Fig. 6 shows that the mass transition appears to be reasonably bracketed at $M_{\rm tr,\,s} \simeq 10^{10.5\pm0.25}\,{\rm M}_\odot$. The mean values of the PDF $1 + \xi$ at $\cos\theta = 0.9$ are, respectively, 0.98 and 1.02 for $M_{\rm s} = 10^{10.75}$ and $10^{10.25}\,{\rm M}_\odot$.

The definition of Hubble type relies on different tracers. Hence, it is of interest to quantify the alignment or misalignment of galaxies classified according to these tracers. One should keep in mind that these tracers are not independent from one another (as illustrated in Appendix A). Top row, middle column of Fig. 5 shows the excess probability of alignment for V/σ. Dispersion-dominated galaxies with small V/σ ratios (i.e. elliptical galaxies) have their spin perpendicular to filaments, while centrifugally supported galaxies with large V/σ (i.e. disc galaxies) have their spins parallel to filaments. The transition between parallel and perpendicular alignment occurs at $V/\sigma = 0.6$. A similar signal, not represented here, is found for intrinsic (three-dimensional) kinematics. The top-right panel of Fig. 5 shows ξ as a function of the sSFR of galaxies. Intense star-forming galaxies that rejuvenate their stellar mass content in less than $1/{\rm sSFR} = 1/10^{-0.5} \simeq 3\,{\rm Gyr}$ tend to align with filaments. Conversely, galaxies that are passive (${\rm sSFR} \simeq 0.1\,{\rm Gyr}^{-1}$) show a random orientation of their spin relative to the filaments. The left and central panels of the middle row of Fig. 5 show ξ as a function of the $g - r$ and the $r - i$ colours of galaxies, respectively. Redder galaxies ($g - r > 0.25$ or $r - i > 0.13$) have their spin perpendicular to their filaments, while bluer galaxies ($g - r \leq 0.25$ or $r - i \leq 0.13$) have their spin parallel to them. The right-hand panel of the middle row of Fig. 5 shows ξ as a function of the stellar metallicity Z. Metal-poor galaxies are more aligned with filaments than metal-rich galaxies which tend to be misaligned. The bottom-left

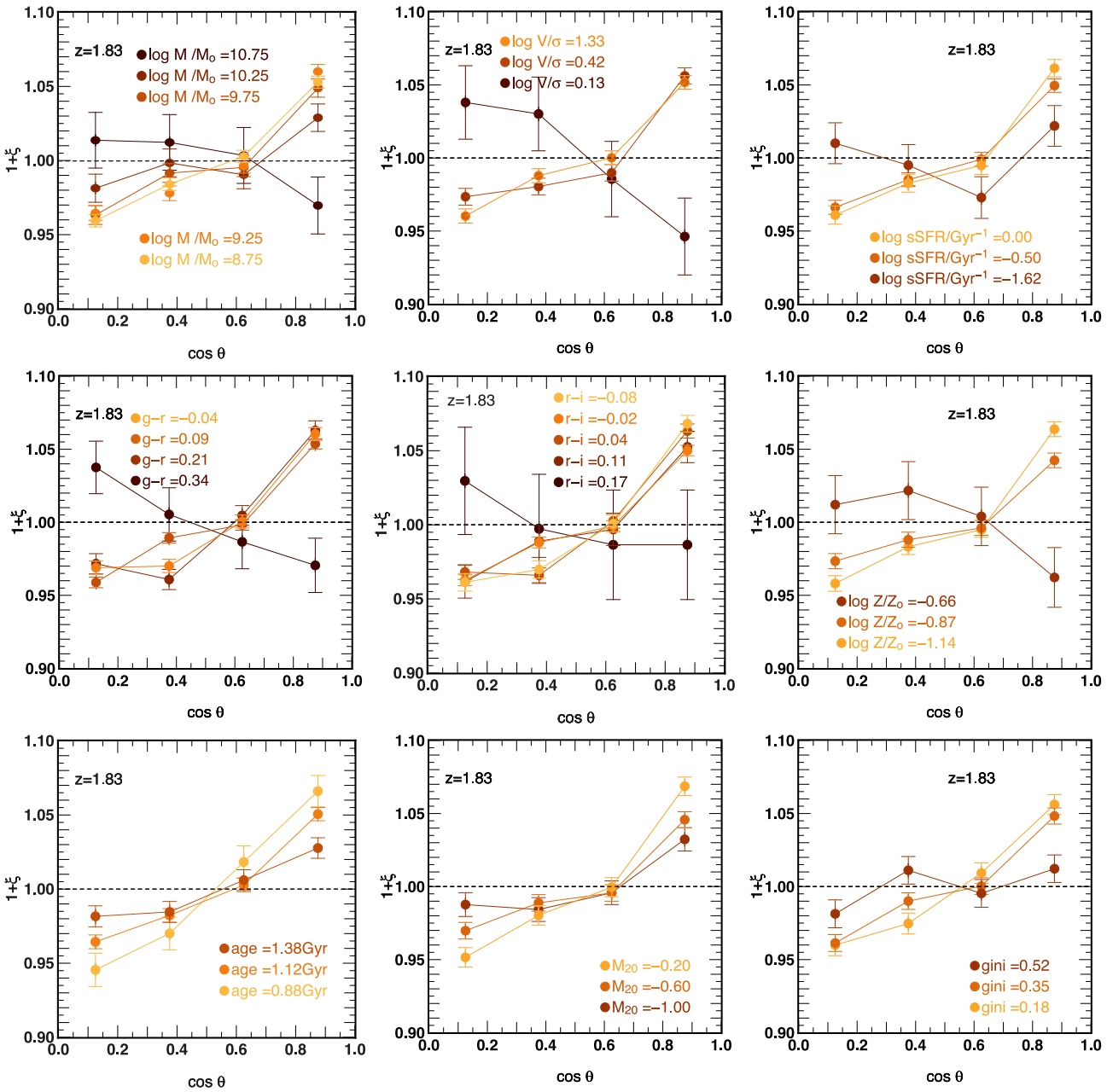


Figure 5. Excess probability, ξ, of the alignment between the spin of galaxies and their closest filament is shown as a function of galaxy properties at $z = 1.83$: M_s (top row, left column), V/σ (top row, middle column), sSFR (top row, right column), $g - r$ (middle row, left column), $r - i$ (middle row, middle column), metallicity Z (middle row, right column), age (bottom row, left column), M_{20} (bottom row, middle column) and Gini (bottom row, right column). Half-sigma error bars are shown for readability. Dashed line is uniform PDF (excess probability $\xi = 0$). Massive, dispersion-dominated, passive, red, smooth and old galaxies tend to have a spin perpendicular or randomly oriented with the direction of their filament. Low-mass, centrifugally supported, star-forming, blue, irregular and young galaxies tend to align with the direction of their closest filament.

panel of Fig. 5 shows ξ as a function of the galaxy age. Older galaxies have their spin more randomly oriented with that of the filaments, and young galaxies with age below $\simeq$1.2 Gyr exhibit a stronger alignment. Finally, the bottom-middle and bottom-right panels of Fig. 5 show ξ as a function of the M_{20} and Gini quantitative morphological indices. Galaxies with high M_{20} are more aligned with filaments than galaxies with low M_{20}. Galaxies with low Gini are more aligned with filaments than galaxies with high Gini. Galaxies with low M_{20} and high Gini tend to trace elliptical galaxies (Lotz et al. 2008). Note that for age, M_{20} and Gini, galaxies in the Horizon-AGN simulation do not seem to present enough leverage to identify a complete misalignment (in contrast to the other tracers).

To summarize, massive, dispersion-dominated, passive, red, smooth, metal-rich and old galaxies tend to have a spin perpendicular or randomly oriented to filaments. In contrast, low-mass, centrifugally supported, star-forming, blue, irregular, metal-poor and young galaxies tend to align with filaments.

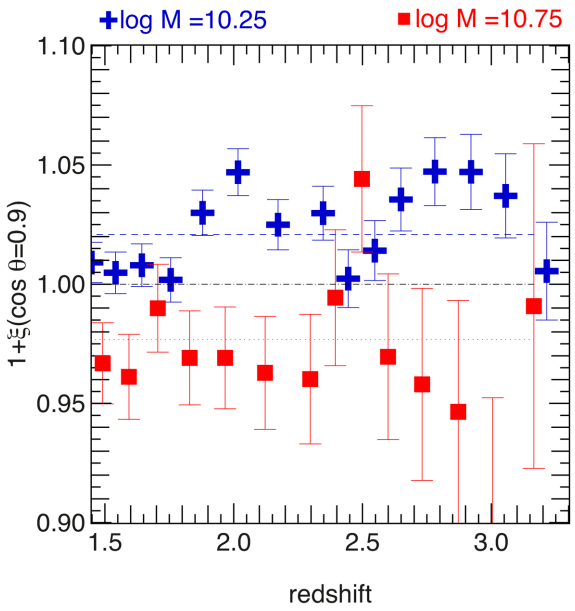


Figure 6. Average values of $1 + \xi(\cos\theta = 0.9)$ as a function of redshift for two different bins of stellar mass $M_s = 10^{10.25}$ (pluses) and $M_s = 10^{10.75}$ (squares). Errors bars correspond to half-sigma error. The dotted (respectively dashed) lines correspond to the mean of the lower (respectively higher) bin mass. The mean values of $1 + \xi(\cos\theta = 0.9)$ are 0.975 for stellar mass $M_s = 10^{10.75}$ and 1.023 for $M_s = 10^{10.25}$, respectively. The transition mass seems reasonably bracketed at $M_{tr,\,s} = 10^{10.5 \pm 0.25}\,M_\odot$.

The transitions presented here are more indicative of a trend than a definite proof that each tracer yields a precise morphological transition. Amongst the various tracers, V/σ and $g - r$ are those for which the transition from alignment to perpendicular misalignment is the most significant. Yet the ensemble allows us to have confidence in the underlying physical picture as they are all consistent with the expected variations. The above-mentioned consistent analysis of its redshift evolution (Fig. 6) brings further confidence in our results. It should also be noted that each estimator is derived from a fairly crude automated analysis.

3.2 Spin orientation along the cosmic web

We now investigate the orientation of the alignment as a function of the distance to filaments and nodes. The upper panel of Fig. 7 shows the evolution of alignment of the spin of galaxies as a function of distance to the closest filament for a low-mass subsample. We apply this measurement to low-mass galaxies because they lie in filaments, sheets and voids, while the most massive galaxies are usually located at the intersection of the most massive filaments in the most massive haloes, therefore, with a minimum scatter in distance to filaments. Galaxies closer to filaments have their spin more parallel. The lower panel of Fig. 7 shows the evolution of alignment of the spin of galaxies as a function of distance to nodes (i.e. where filaments intersect) along the filaments. Galaxies further away from nodes have their spin more parallel than galaxies closer to nodes. This is consistent with the idea that galaxies merge while drifting along filaments (which destroys alignment), and with the strong colour (curvilinear) gradients found by Gay et al. (2010).

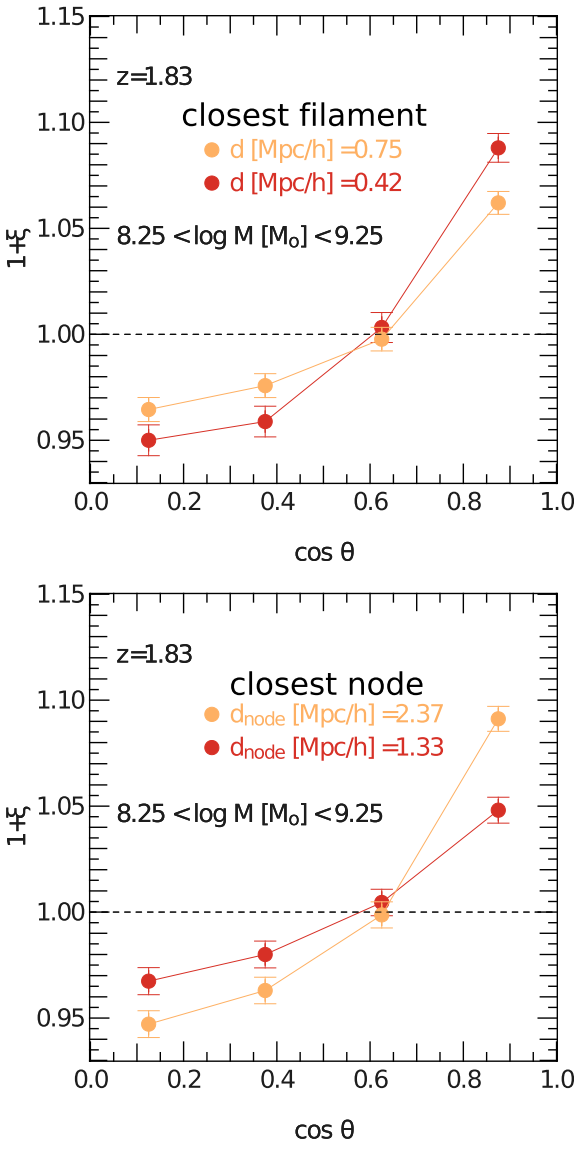


Figure 7. Excess probability ξ of the alignment between the spin of galaxies and their closest filament as a function of their distance to the closest filament (top panel) or node (bottom panel) is shown at $z = 1.83$. Galaxies closer to filaments have their spin more parallel, while galaxies closer to nodes are more randomly oriented. Dashed line is the zero excess probability $\xi = 0$.

3.3 Redshift evolution

We now investigate the redshift evolution of the excess probability of alignment. We post-process the Horizon-AGN in the following redshift range: $z = 3.01$–1.23. Fig. 8 shows the amplitude of the alignment of the spin of all galaxies as a function of redshift. The PDF shows that, on average, galaxies are aligned with their neighbouring filament, because low-mass galaxies dominate in number over massive galaxies (because the mass function of galaxies is strongly decreasing with mass). The amplitude of the alignment decreases with cosmic time (decreasing redshift) which is a result of more galaxies evolving passively (i.e. for a given mass, the SFR decreases with time). The lower the redshift, the stronger the amount of shell crossing and cumulative contribution from mergers along the filaments which tend to destroy the existing alignment and

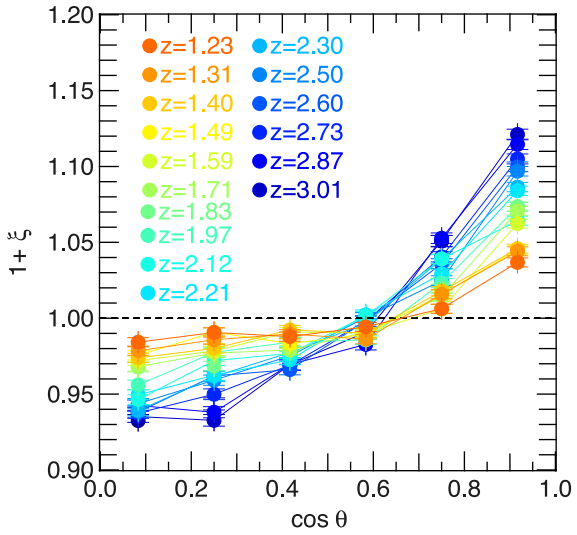


Figure 8. Excess probability ξ of the alignment between the spin of galaxies (with mass above 10^8 $M_\odot$) and their closest filament as a function of redshift is shown. Dashed line is the zero excess probability $\xi = 0$. The amplitude of the correlation shows an alignment which increases with redshift (i.e. decreases with cosmic time).

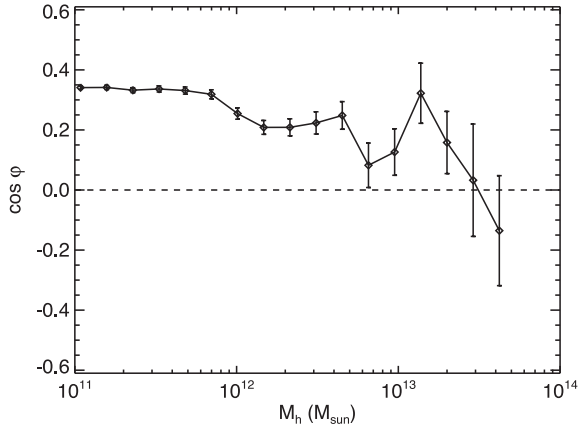


Figure 9. Average cosine of the angle $\cos\psi$ between the galaxy angular momentum and that of its host halo as a function of the halo mass at $z = 1.3$ (solid line). Error bars are the standard errors of the mean. Note that the random distribution (dashed line) is for $\cos\psi = 0$ because vectors can be pointing towards different directions and, therefore, galaxy spin and halo spin can be aligned or anti-aligned.

convert orbital momentum into spin to orient it perpendicular to that of the host filament.

4 DISCUSSION: AGN FEEDBACK PROMOTES SPIN SWINGS?

The above analysis suggests that the most massive galaxies, which are also the reddest, the oldest, the most metal-rich and the most pressure-supported, tend to have their spin perpendicular to the axis of filaments. Galaxies on average also show less alignment with time (i.e. with decreasing redshift). We argue that the origin of the misalignment is the sudden reorientation of galactic angular momentum during mergers. This was shown in Codis et al. (2012) to be the case for the origin of the DM halo–filament misalignment for massive haloes. Misalignment occurs because orbital momentum is converted into spin, as DM haloes catch up each other *along* the filaments. We also argue that indirect merger rate indicators, such as those presented in Section 2.2 (galaxy properties), can only be modelled once an efficient feedback mechanism is implemented to produce morphological and physical diversity.

4.1 The contribution of cosmic dynamics

Massive galaxies undergo mergers (minor and major) that contribute to misalignment of their spin with respect to the direction of the filament. Given the level of significance for this range of galaxy mass (compatible with a uniform PDF at less than 1σ for $M_s \simeq 10^{10.75}$ $M_\odot$, see the top-left panel of Fig. 5), it is still unclear whether massive galaxies have as strong a preference for a spin orientation perpendicular to the axis of their neighbouring filament as do their halo counterparts, or if they tend to be randomly oriented.

First, note that, globally, the excess probability of alignment is always at values $\xi \lesssim 0.1$ (Figs 5, 7 and 8); therefore, the signal is weak (not all galaxies align or misalign) but statistically significant. For instance, the (mis)alignment signal between haloes and filaments (Codis et al. 2012) or that between the large-scale vorticity and filaments (Laigle et al. 2013) are respectively of somewhat (∼15 per cent) and significantly (∼100 per cent) larger amplitudes. One could expect that small scales (i.e. galaxies) decouple more strongly from the large-scale filaments than the intermediate scales (i.e. haloes), which are the first virialized structures. In particular, there could be a significant amount of redistribution of angular momentum within the inner regions of haloes (Kimm et al. 2011; Danovich et al. 2012; Dubois et al. 2012b; Kassin et al. 2012a,b; Tillson et al. 2012), and as a consequence, galaxies misalign with the spin of their host halo (see Fig. 9 for the average cosine of the angle ψ between the angular momentum of the galaxy and that of the host halo), while they keep a strong alignment with the spin of the dark halo's central region (Hahn et al. 2010). On the other hand, low-mass central galaxies are fed angular momentum directly by cold flows (Pichon et al. 2011; Tillson et al. 2012), which connects them more tightly to their cosmic environment than one would naively expect. The orientation and amplitude of the stellar component itself reflect the corresponding cumulative advection of cold gas directly on to the circumgalactic medium. In contrast, the orientation of the spin of haloes is more sensitive to the latest stochastic accretion events at the virial radius. The net outcome of both competing processes, as traced by the physical properties of galaxies, is summarized in Figs 5 and 6.

Fig. 9 shows that, because of the above-mentioned redistribution of angular momentum within the inner region of the halo, the galactic spin is weakly correlated to that of the whole halo, and the effect is more pronounced for more massive haloes which merge more frequently. Satellites end up reaching the central galaxy with less correlated orbital angular momentum even though they globally originate from a preferred direction, as set by the cosmic web. In order to test this hypothesis, we build merger trees from the catalogue of galactic structures detected by our galaxy finder. For each galaxy, we measure the stellar mass acquired through the different branches of the tree (satellites) that we quote as a merger, the main progenitor being excluded from the calculation. Fig. 10 shows that massive galaxies acquire a non-negligible fraction of their mass by mergers (at least 1000 particles of star particles, up to 20 per cent at $z = 1.83$), while low-mass galaxies grow their stellar mass content almost exclusively by *in situ* star formation (e.g. De Lucia & Blaizot 2007; Oser et al. 2010). Fig. 11 shows examples of the evolution of the spin orientation for six massive galaxies,

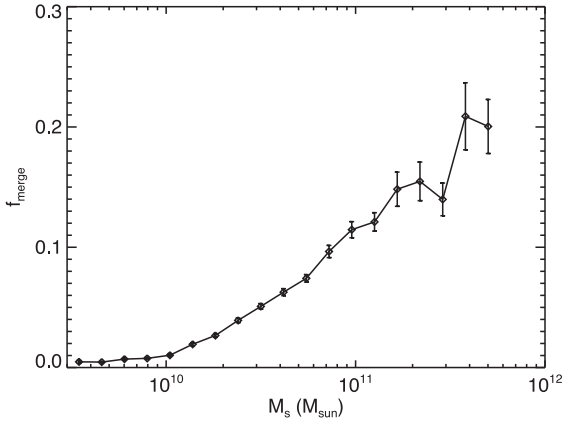


Figure 10. Average fraction of stellar mass gained through mergers as a function of the galaxy stellar mass at $z = 1.83$. The error bars are the standard errors on the mean. More massive galaxies have a larger fraction of galaxy mergers contributing to their stellar mass. Lower mass galaxies build up their stellar mass through *in situ* star formation only.

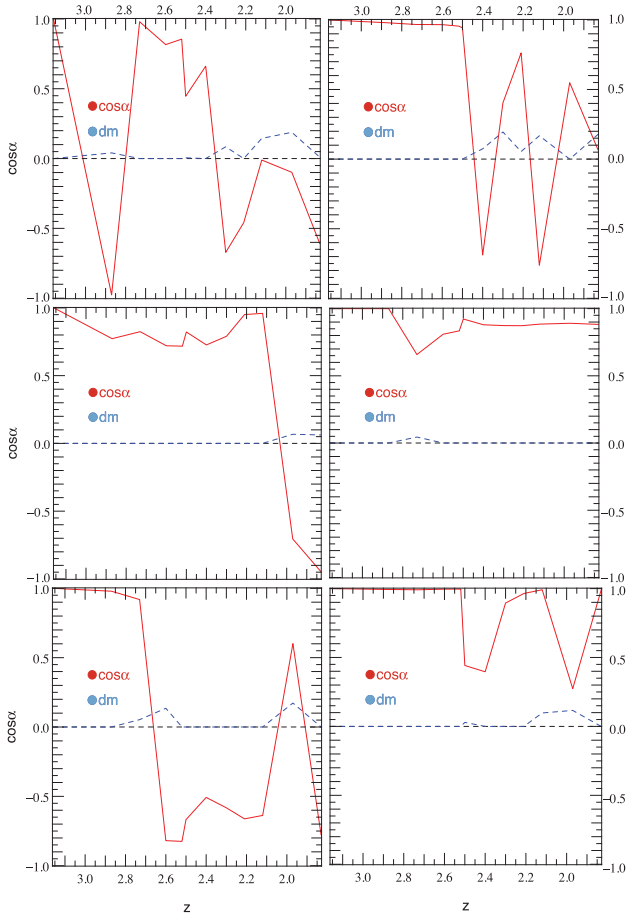


Figure 11. Examples of galaxies changing their spin direction during mergers with stellar mass $1.7 \times 10^{11}\,M_\odot$ (top left), $7.3 \times 10^{10}\,M_\odot$ (top right), $3.8 \times 10^{10}\,M_\odot$ (middle left), $4.8 \times 10^{10}\,M_\odot$ (middle right), $1.2 \times 10^{11}\,M_\odot$ (bottom left) and $6.0 \times 10^{10}\,M_\odot$ (bottom right) at $z = 1.83$. $\cos\alpha$ (red curve) is the cosine of the angle between the spin of the galaxy at the current redshift and the initial spin measured at $z = 3.15$. The differential fraction of mass between two time steps coming from mergers $\mathrm{d}m = \delta M_{\rm merge}/M_{\rm s}$ (in blue) is overplotted. Non-zero values correspond to rapid changes in spin direction. In the absence of mergers, the galaxy spin has a steady direction.

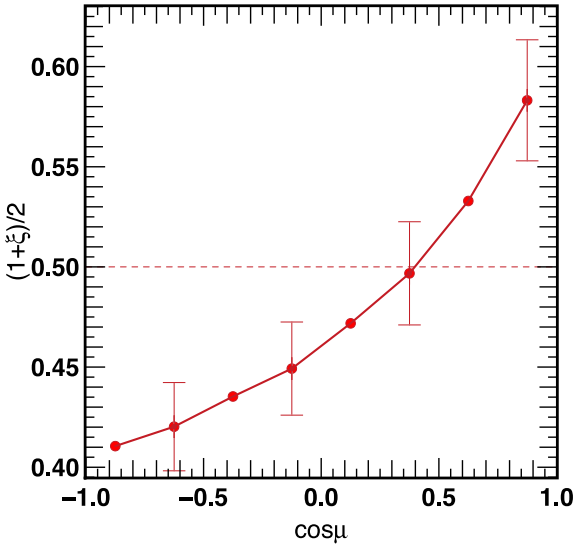


Figure 12. Excess probability ξ of the cosine of the angle μ between the vorticity of the gas and the direction of the spin of the galaxies at $z = 1.83$. The dashed line indicates no correlation. Vorticity is computed as the curl of the velocity field, after a Gaussian smoothing of the velocity field with a kernel length of $780\,h^{-1}$ kpc. Note that fewer galaxies are anti-aligned. Error bars are the standard errors of the mean.

$4 \times 10^{10} \lesssim M_{\rm s} \lesssim 2 \times 10^{11}\,M_\odot$. They have a significant contribution from mergers to their stellar mass, which play a significant role in shaping their spin orientations (Bett & Frenk 2012). In Fig. 11, the fraction of mass gained by mergers $\delta M_{\rm merge}/M_{\rm s}$ between two time steps is indicated by dashed blue lines. When no mergers happen, galaxies keep a steady spin direction. It is only when a companion galaxy is captured ($\delta M_{\rm merge}/M_{\rm s} \neq 0$) do we see a sudden reorientation of the spin. An investigation of the relative role of minor, major, dry and wet mergers is postponed to a companion paper (Welker et al. 2014) which shows unambiguously that major mergers are indeed responsible for important spin swings.

In contrast to high-mass galaxies, low-mass galaxies have their spins preferentially aligned with that of their closest filaments. Gas embedded within large-scale walls streams into the filaments which bound them, winding up to form the first generation of galaxies which have spins parallel to these filaments (Pichon et al. 2011). Since these galaxies build up their stellar mass *in situ* without significant external perturbations, the stars retain the angular momentum of the cold gas obtained directly from the cosmic web. Fig. 12 shows the excess probability of the cosine of the angle μ between the vorticity of the gas (as estimated on scales of $200\,h^{-1}$ kpc) at the galaxy's position and the direction of the spin of galaxies (dominated by the low-mass population). As was found in Laigle et al. (2013) for the spin of DM haloes (see also Libeskind et al. 2013b), the galactic spin is also strongly correlated with the vorticity of the surrounding gas. This correlation has polarity: there are fewer galaxies with their spin anti-aligned with the local vorticity. This dynamical and stellar evidence therefore allows us to apply to baryons the scenario presented in Laigle et al. (2013) on the vorticity-driven origin of the galactic spin–filament alignment.

4.2 The contribution of AGN feedback

Fig. 13 shows the stellar halo mass relation at $z = 1.3$ for the Horizon-AGN simulation. It appears that above $M_{\rm h} \gtrsim 10^{12}\,M_\odot$

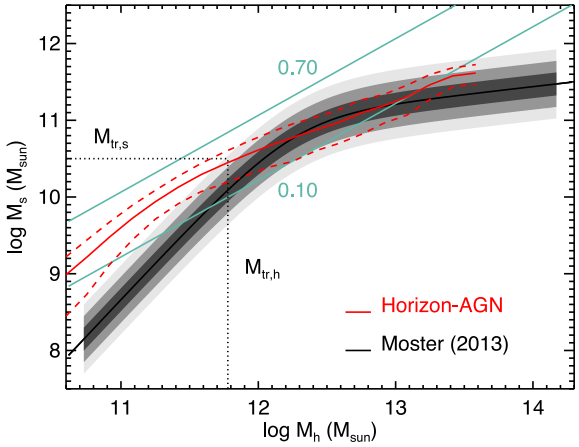


Figure 13. Average stellar mass of the central galaxy as a function of the host halo mass in the simulation at $z = 1.3$ (red curve) with the 1σ dispersion together with the abundance matching result from Moster et al. (2013) at the same redshift assuming a Salpeter IMF (+0.19 dex in stellar mass from Kroupa to Salpeter IMF) with the 1σ, 2σ and 3σ error bars represented by the shaded areas. The cyan lines indicate constant stellar efficiencies defined as $f_{\rm eff} = M_{\rm s}/(\Omega_{\rm b}/\Omega_{\rm m}M_{\rm h})$. The transition mass between alignment and misalignment of the galaxy ($M_{\rm tr,\,s}$), respectively, halo ($M_{\rm tr,\,h}$), spin and the filament are represented as dotted lines.

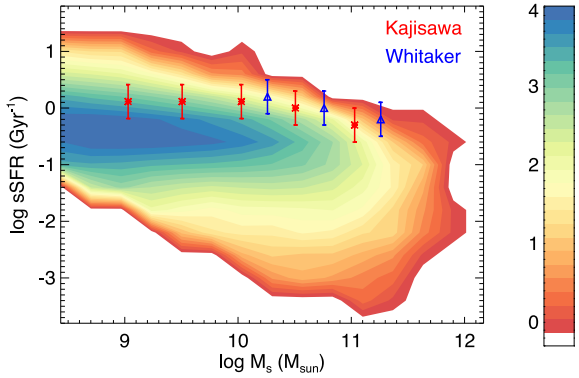


Figure 14. Diagram of the distribution of galaxies as a function of their sSFR and stellar mass $M_{\rm s}$ at $z = 1.3$. The contours are the number of galaxies in log units. Red and blue points with error bars are the observations at $z = 1.25$ extracted from Behroozi, Wechsler & Conroy (2013) and originally from Kajisawa et al. (2010) and Whitaker et al. (2012) and rescaled for a Salpeter IMF (+0.26 dex in stellar mass from Chabrier to Salpeter IMF). The sSFR decreases with galaxy stellar mass.

the relation is in good agreement with abundance-matching results from Moster et al. (2013). Note that the stellar masses from Moster et al. (2013) are rescaled by +0.19 dex to account for a change from Kroupa to Salpeter IMF. The presence of the feedback from AGN reduces the amount of stars formed in these massive galaxies, allowing them to agree with observations by reducing the stellar mass up to one order of magnitude (Dubois et al. 2013; Puchwein & Springel 2013; Martizzi et al. 2014). Note also that the transition mass from alignment to misalignment is close to the mass ($\simeq 10^{11}$ M$_\odot$) where passive galaxies become more abundant than star-forming galaxies in observations (Drory et al. 2009; Davidzon et al. 2013; Ilbert et al. 2013). For massive galaxies, mergers are key in both swinging the spin *and* changing the morphology. The sSFR decreases significantly with galaxy stellar mass (see Fig. 14) because of the quenching of their gas accretion rate – in the shock-heated mode of accretion for massive haloes (e.g. Birnboim & Dekel 2003; Ocvirk, Pichon & Teyssier 2008; Dekel et al. 2009) – and because of the strong suppression of cold gas within galaxies via AGN feedback. Observational data from Behroozi et al. (2013) at $z = 1.3$ (originally from Kajisawa et al. 2010 and Whitaker et al. 2012) are represented on top of the distribution of our simulations points. Stellar masses from Behroozi et al. (2013) are rescaled by +0.26 dex to account for a change from Chabrier to Salpeter IMF. The quenching of the SFR in massive galaxies through AGN feedback leads to an enhanced fraction of stars gained through mergers (Dubois et al. 2013). For lower mass galaxies, the agreement of the stellar halo mass relation with observational data is less favourable because feedback from SNe is not strong enough to suppress the star formation in dwarfs. Some missing physical processes, such as radiation from young stars, are probably necessary to further suppress the star formation in low-mass galaxies (e.g. Hopkins, Quataert & Murray 2011; Murray, Ménard & Thompson 2011; Hopkins et al. 2013). Vogelsberger et al. (2013) manage to reproduce the low-mass tail of the stellar-to-halo mass relation by decoupling hydrodynamically the launched wind gas from the dense star-forming gas (as introduced by Springel & Hernquist 2003). This decoupling of gas is known to generate a more efficient transfer of energy from SNe to large-scale galactic winds, compared to local prescriptions as we have adopted here (as shown by Dalla Vecchia & Schaye 2008), but it lacks physical motivation. Note that the choice of the Salpeter IMF instead of a Chabrier IMF decreases the total energy released by a stellar particle by a factor of 3 (assuming a minimum and maximum mass of 0.1 and 100 M$_\odot$). However, we are still confident that stronger feedback in low-mass galaxies should not drastically change their orientation, since mergers for that class of haloes are rare enough (Fakhouri, Ma & Boylan-Kolchin 2010; Genel et al. 2010). This is in particular true for the major mergers that are required to significantly reorient the spins of galaxies (Welker et al. 2014).

The significant contribution from AGN feedback at reducing the stellar activity in massive galaxies is mandatory to obtain a diversity in the physical properties of galaxies (colours, V/σ, sSFR, etc.) across the whole mass range (see Appendix A). The effect of AGN feedback is to also reduce the mass of stars formed *in situ*, i.e. to prevent the formation of a rotation-supported component in massive galaxies and to turn massive galaxies into pressure-supported ellipticals (Dubois et al. 2013). In the absence of a central supermassive BH, the magnitude of the angular momentum of the stellar component of massive galaxies will therefore be larger, as a larger fraction of their (larger) stellar mass will be distributed in a rotationally supported disc. Thus, a merging satellite the angular momentum of which is misaligned with that of the central galaxy produces a variation in the angle between the pre-merger and the post-merger spin of the galaxy that is smaller for the disc case (no AGN case) than for the elliptical case (AGN case). Moreover, for massive galaxies, the feedback from the central AGN switches off later accretion of circumgalactic gas (Dubois et al. 2010). Consequently, the possible realignment of the galactic spin with the filament that could operate after a merger due the accretion of fresh gas is reduced by the presence of the AGN feedback from the galaxy remnant. AGN feedback thereby acts as a catalyst of spin swings.

5 CONCLUSIONS

Our analysis of the Horizon-AGN flagship simulation, which models AGN as well as stellar feedback so as to produce morphological

diversity, shows that the orientation of the spin of galaxies depends on various galaxy properties such as stellar mass, V/σ, sSFR, M_{20}, metallicity, colour and age. The spins of galaxies tend to be preferentially parallel to their neighbouring filaments, for low-mass, young, centrifugally supported, metal-poor, bluer galaxies and perpendicular for higher mass, higher velocity dispersion, red, metal-rich old galaxies. The alignment is the strongest, the closer to the filaments and further from the nodes of the cosmic web the galaxies are. This is in agreement with the predictions of Codis et al. (2012) for DM haloes. We find a transition mass, $M_{\rm tr,\,s} \simeq 3 \times 10^{10}\,{\rm M}_{\odot}$ which is also consistent with these authors' predictions for the corresponding halo transition mass. Due to the weak galaxy–halo alignment, the amplitude of the correlation with cosmic filaments is somewhat weaker for galaxies than for haloes. It also decreases with cosmic time due to mergers and quenching of cold flows and star formation. Hence, our results suggest that galaxy properties can be used to trace the spin swings along the cosmic web.

The transition from the aligned to the misaligned case is dynamically triggered by mergers (the frequency of which increases with galaxy mass) that swing the spin of galaxies. AGN feedback has a key role at preventing further gas inflow and quenching the *in situ* star formation after such an event, in order to preserve the misalignment operated by the merger.

Though it is expected that galaxy modelling will improve over the next decade – in particular the way feedback is taken into account in large-scale cosmological simulations – we do not anticipate that the particular results found in this paper should change qualitatively. The finding that the morphological diversity of galaxies traces populations with different spin–filament alignments, which in turn is in part inherited from the anisotropy of the embedding cosmic web, is both robust predictions of the current gravitational clustering scenario and of our understanding of the dynamics of elliptical and spiral galaxies.

In a forthcoming paper, we will analyse more realistic mock catalogues from the Horizon-AGN light-cone down to a lower redshift to investigate the amount of modification of the signal induced by dust extinction, projection effects, limited resolution and finite signal-to-noise ratio. More efficient and robust estimators for morphology, either intrinsic using the full data set of the simulation or projected using virtual degraded observables, will be built and compared. Quantitative comparisons to observations are postponed to this paper.

ACKNOWLEDGEMENTS

We thank the anonymous referee for suggestions, which improved the clarity of the paper. This work has made use of the HPC resources of CINES (Jade supercomputer) under the allocation 2013047012 made by GENCI. The post-processing made use of the `horizon` and `Dirac` clusters. This work is partially supported by the Spin(e) grants ANR-13-BS05-0002 of the French *Agence Nationale de la Recherche* and by the National Science Foundation under Grant No. NSF PHY11-25915. This research is part of the Horizon-UK project. YD and CP thank the KITP for hospitality when this project was initiated. CP also thanks the institute of Astronomy for a Sacker visiting fellowship and the PEPS 'Physique théorique et ses interfaces' for funding. The research of YD and JS has been supported at IAP by ERC project 267117 (DARK) hosted by Université Pierre et Marie Curie – Paris 6. YD is grateful to the Beecroft Institute for Particle Astrophysics and Cosmology for hospitality and to the Balzan Foundation for support, at Oxford University where some of this work was carried out. The research of JS has also been supported at the Johns Hopkins University by National Science Foundation grant OIA-1124403. The research of AS and JD at Oxford is supported by the Oxford Martin School and Adrian Beecroft. This work was also supported by the ILP LABEX (under reference ANR-10-LABX-63) was supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02. We thank D. Munro for freely distributing his YORICK programming language and opengl interface (available at http://yorick.sourceforge.net/).

REFERENCES

Abraham R. G., Valdes F., Yee H. K. C., van den Bergh S., 1994, ApJ, 432, 75
Abraham R. G., van den Bergh S., Nair P., 2003, ApJ, 588, 218
Abraham R. G. et al., 2007, ApJ, 669, 184
Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007, ApJ, 655, L5
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bailin J., Steinmetz M., 2005, ApJ, 627
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bett P. E., Frenk C. S., 2012, MNRAS, 420, 3324
Birnboim Y., Dekel A., 2003, MNRAS, 345, 349
Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320
Colless M. et al., 2001, MNRAS, 328, 1039
Croft R. A. C., Metzler C. A., 2000, ApJ, 545, 561
Croton D. J. et al., 2006, MNRAS, 365, 11
Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431
Danovich M., Dekel A., Hahn O., Teyssier R., 2012, MNRAS, 422, 1732
Davidzon I. et al., 2013, A&A, 558, A23
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
de Lapparent V., Geller M. J., Huchra J. P., 1986, ApJ, 302, L1
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dekel A. et al., 2009, Nature, 457, 451
Devriendt J. et al., 2010, MNRAS, 403, L84
Doroshkevich A. G., 1970, Astrophysics, 6, 320
Drory N. et al., 2009, ApJ, 707, 1595
Dubois Y., Teyssier R., 2008, A&A, 477, 79
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2010, MNRAS, 409, 985
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012a, MNRAS, 420, 2662
Dubois Y., Pichon C., Haehnelt M., Kimm T., Slyz A., Devriendt J., Pogosyan D., 2012b, MNRAS, 423, 3616
Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, MNRAS, 433, 3297
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Faltenbacher A., Gottlöber S., Kerscher M., Müller V., 2002, A&A, 395, 1
Gay C., Pichon C., Le Borgne D., Teyssier R., Sousbie T., Devriendt J., 2010, MNRAS, 404, 1801
Geller M. J., Huchra J. P., 1989, Science, 246, 897
Genel S., Bouché N., Naab T., Sternberg A., Genzel R., 2010, ApJ, 719, 229
Greggio L., Renzini A., 1983, A&A, 118, 217
Haardt F., Madau P., 1996, ApJ, 461, 20
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007a, MNRAS, 375, 489
Hahn O., Carollo C. M., Porciani C., Dekel A., 2007b, MNRAS, 381, 41
Hahn O., Teyssier R., Carollo C. M., 2010, MNRAS, 405, 274
Hatton S., Ninin S., 2001, MNRAS, 322, 576
Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526

Hockney R. W., Eastwood J. W., 1981, Computer Simulation Using Particles. McGraw-Hill, New York
Hopkins P. F., Quataert E., Murray N., 2011, MNRAS, 417, 950
Hopkins P. F., Keres D., Onorbe J., Faucher-Giguere C.-A., Quataert E., Murray N., Bullock J. S., 2013, preprint (arXiv:e-prints)
Hoyle F., 1949, Problems of Cosmical Aerodynamics. Central Air Documents Office, Dayton, OH, p. 195
Huchra J., Davis M., Latham D., Tonry J., 1983, ApJS, 52, 89
Ilbert O. et al., 2013, A&A, 556, A55
Joachimi B., Semboloni E., Bett P. E., Hartlap J., Hilbert S., Hoekstra H., Schneider P., Schrabback T., 2013, MNRAS, 431, 477
Kajisawa M., Ichikawa T., Yamada T., Uchimoto Y. K., Yoshikawa T., Akiyama M., Onodera M., 2010, ApJ, 723, 129
Kassin S. A., Devriendt J., Fall S. M., de Jong R. S., Allgood B., Primack J. R., 2012a, MNRAS, 424, 502
Kassin S. A. et al., 2012b, ApJ, 758, 106
Kennicutt R. C., Jr, 1998, ApJ, 498, 541
Kimm T., Devriendt J., Slyz A., Pichon C., Kassin S. A., Dubois Y., 2011, preprint (arXiv:e-prints)
Klypin A. A., Shandarin S. F., 1983, MNRAS, 204, 891
Komatsu E. et al., 2011, ApJS, 192, 18
Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304
Laigle C. et al., 2013, preprint (arXiv:e-prints)
Lee J., Pen U., 2000, ApJ, 532, L5
Lee J., Pen U.-L., 2002, ApJ, 567, L111
Lee B. et al., 2013, ApJ, 774, 47
Leitherer C. et al., 1999, ApJS, 123, 3
Leitherer C., Ortiz Otálvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, ApJS, 189, 309
Libeskind N. I., Hoffman Y., Knebe A., Steinmetz M., Gottlöber S., Metuki O., Yepes G., 2012, MNRAS, 421, L137
Libeskind N. I., Hoffman Y., Forero-Romero J., Gottlöber S., Knebe A., Steinmetz M., Klypin A., 2013a, MNRAS, 428, 2489
Libeskind N. I., Hoffman Y., Steinmetz M., Gottlöber S., Knebe A., Hess S., 2013b, ApJ, 766, L15
Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163
Lotz J. M. et al., 2008, ApJ, 672, 177
Martizzi D., Mohammed I., Teyssier R., Moore B., 2014, MNRAS, 440, 2290
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Murray N., Ménard B., Thompson T. A., 2011, ApJ, 735, 66
Ocvirk P., Pichon C., Teyssier R., 2008, MNRAS, 390, 1326
Oesch P. A. et al., 2010, ApJ, 714, L47
Omma H., Binney J., Bryan G., Slyz A., 2004, MNRAS, 348, 1105
Oser L., Ostriker J. P., Naab T., Johansson P. H., Burkert A., 2010, ApJ, 725, 2312
Paz D. J., Stasyszyn F., Padilla N. D., 2008, MNRAS, 389, 1127
Peebles P. J. E., 1969, ApJ, 155, 393
Pichon C., Pogosyan D., Kimm T., Slyz A., Devriendt J., Dubois Y., 2011, MNRAS, 1739
Planck Collaboration, 2013, preprint (arXiv:e-prints)
Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R., Gottloeber S., 2008, ApJS, 178, 179
Puchwein E., Springel V., 2013, MNRAS, 428, 2966
Rasera Y., Teyssier R., 2006, A&A, 445, 1
Salpeter E. E., 1955, ApJ, 121, 161
Schaefer B. M., 2009, Int. J. Mod. Phys. D, 18, 173
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Shandarin S. F., Zeldovich Y. B., 1989, Rev. Mod. Phys., 61, 185
Sousbie T., Pichon C., Colombi S., Pogosyan D., 2008, MNRAS, 383, 1655
Sousbie T., Colombi S., Pichon C., 2009, MNRAS, 393, 457
Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Sugerman B., Summers F. J., Kamionkowski M., 2000, MNRAS, 311, 762
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Tegmark M. et al., 2004, ApJ, 606, 702
Tempel E., Libeskind N. I., 2013, ApJ, 775, L42
Tempel E., Stoica R. S., Saar E., 2013, MNRAS, 428, 1827
Teyssier R., 2002, A&A, 385, 337
Tillson H., Devriendt J., Slyz A., Miller L., Pichon C., 2012, preprint (arXiv:e-prints)
Toro E. F., Spruce M., Speares W., 1994, Shock Waves, 4, 25
Trowland H. E., Lewis G. F., Bland-Hawthorn J., 2013, ApJ, 762, 72
Trujillo I., Carretero C., Patiri S. G., 2006, ApJ, 640, L111
Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647
Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
Wang T. et al., 2012, ApJ, 752, 134
Welker C., Devriendt J., Dubois Y., Pichon C., Peirani S., 2014, in press
Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, ApJ, 754, L29
White S. D. M., 1984, ApJ, 286, 38
Zel'dovich Y. B., 1970, A&A, 5, 84
Zeldovich I. B., Einasto J., Shandarin S. F., 1982, Nature, 300, 407
Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747
Zhang Y., Yang X., Wang H., Wang L., Mo H. J., van den Bosch F. C., 2013, ApJ, 779, 160

APPENDIX A: PROPERTIES OF GALAXIES

Fig. A1 shows various physical properties of galaxies: V/σ, sSFR, $g - r$ colour and age as a function of the stellar mass at $z = 1.83$. There are correlations between the stellar kinematics V/σ and galaxy masses: more massive galaxies are pressure supported; their sSFR and mass: more massive galaxies have lower sSFR, i.e. are more passive; their colour and mass: more massive galaxies are redder; and their age and mass: more massive galaxies are older. Note the quite large scatter at all masses, that is more pronounced for lower mass objects, which can be explained by the fact that low-mass field galaxies in low-density environments evolve differently from low-mass satellites galaxies in high-density environments.

Fig. A2 shows the relation between M_s and the morphology parameters Gini and M_{20} measured in the rest-frame u band. These morphological estimators allow in principle to separate spirals from spheroids, ellipticals and merging galaxies. This has been done in the local universe (Lotz et al. 2004) and at higher redshift (Abraham et al. 2007; Lotz et al. 2008; Lee et al. 2013). The regions drawn in the figure are taken from Lotz et al. (2008). Their precise locations with respect to the distributions measured here should be taken with caution: extinction by dust is not taken into account, morphological k-corrections (although quite small) are not accounted for and the spatial resolution is not matched. Despite all these caveats which may explain the relatively small number of galaxies classified as ellipticals or spheroids from these diagrams, there seems to be, qualitatively, rather good agreement between the distributions measured in the deep surveys and in the simulation.

APPENDIX B: GRID-LOCKING

Fig. B1 displays a Mollweide projection of the orientations of *galaxy* spins along Cartesian axes, for a range of halo mass. Galaxies hosted by haloes lighter or heavier than $5 \times 10^{11}\,\mathrm{M}_\odot$ are considered. While the spins of the less massive galaxies are clearly aligned with the grid, no obvious alignment is seen for the high-mass galaxies. Lighter galaxies are preferentially locked with the grid because they are composed of very few grid elements: the gaseous disc of a galaxy with $\sim 10^9\,\mathrm{M}_\odot$, embedded in a halo of

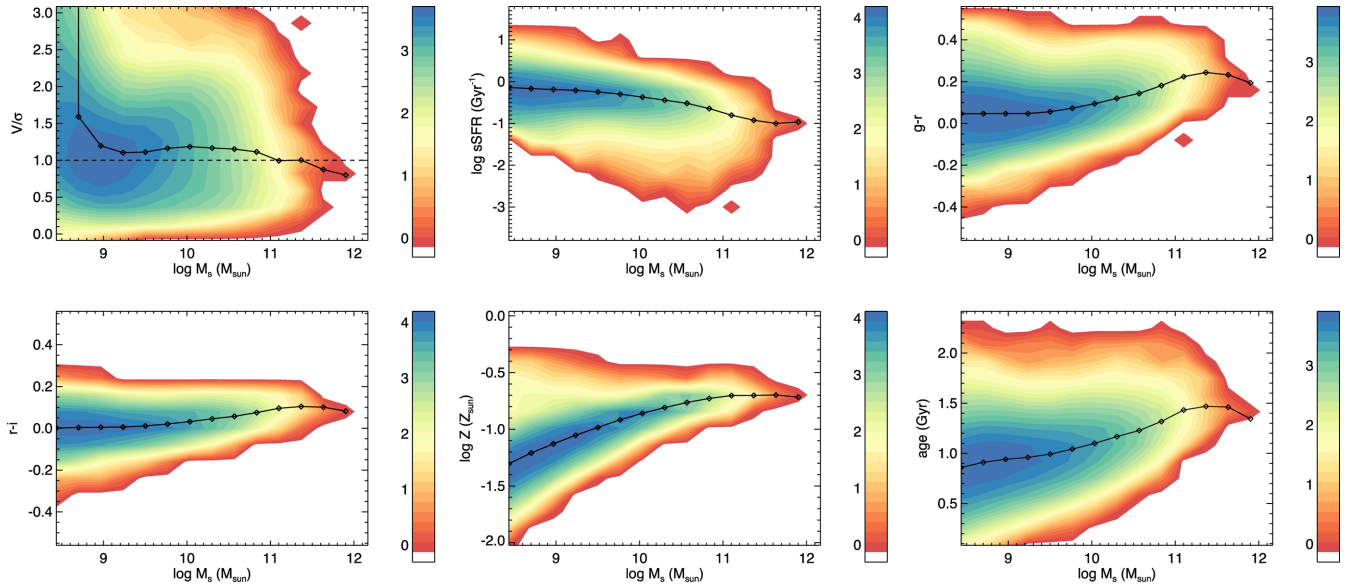


Figure A1. From top to bottom and left to right: contours of the logarithmic number of galaxies for 2D projected kinematics of the stars V/σ, sSFR, $g-r$ colour, $r-i$ colour, metallicity Z and age as a function of the stellar mass of the galaxy at $z = 1.83$. The solid lines with diamonds correspond to the average values as a function of the stellar mass. The dashed line in the top-left panel corresponds to $V/\sigma = 1$.

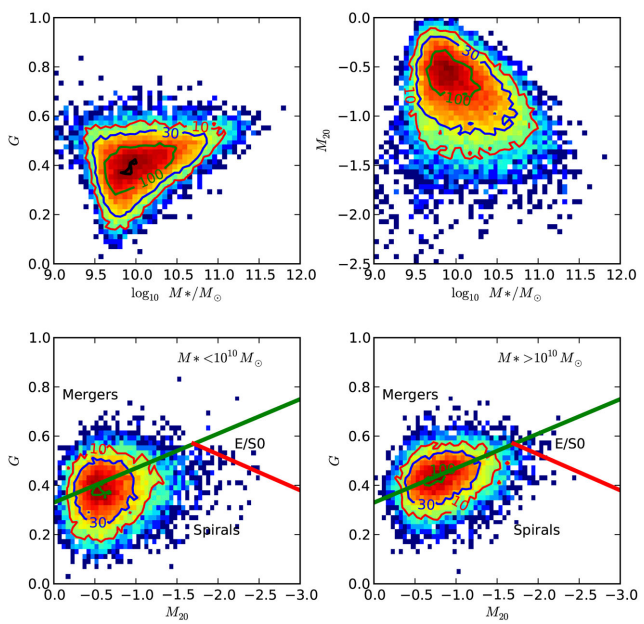


Figure A2. Bivariate distributions of stellar mass, Gini (G) and M_{20} morphological parameters for galaxies in the simulation at $z = 1.8$ in the rest-frame u filter. Lower panels show the Gini–M_{20} distributions for stellar mass selected samples. The zones for elliptical, spiral and merger galaxies are schematically drawn as tentative locations of local ellipticals, spirals and mergers at $z \simeq 2$. See the text for details.

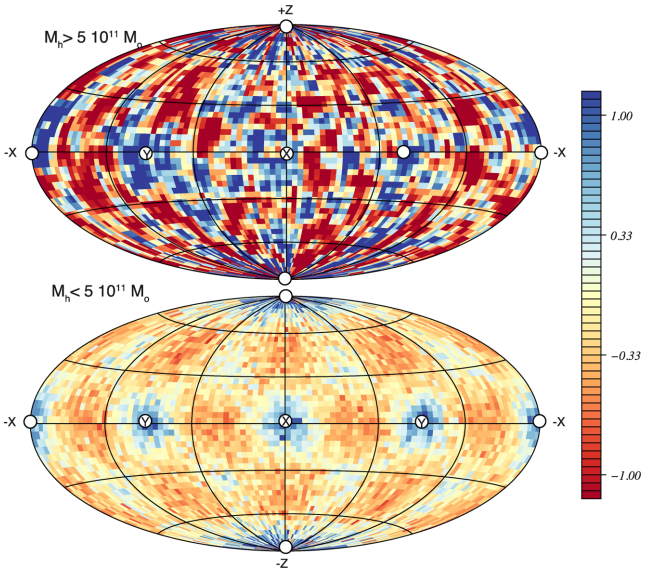


Figure B1. The effect of grid-locking at redshift $z = 1.3$ of the spin of *galaxies* for haloes more (top panel), respectively less (bottom panel) massive than 5×10^{11} M$_\odot$ respectively. The white discs represent the directions of the simulation box as labelled. The colour coding represents relative fluctuations around the mean. The smaller galaxies (bottom panel) show a clear sign of grid-locking, while the more massive sample (top panel) does not.

mass $\sim 10^{11}$ M$_\odot$, tends to be aligned with one of the Cartesian axes due to the anisotropic numerical errors. However, for more massive galaxies, the grid-locking is absent due to a larger number of resolution elements to describe those objects. This result is consistent with that of Hahn et al. (2010) and Danovich et al. (2012).

Fig. B2 shows the distribution of the axis of all the elements of the skeleton on the sphere. There is no preferential direction of alignment with respect to the box axes.

Low-mass galaxies (within halo of mass $<5 \times 10^{11}$ M$_\odot$) show some preferential alignment along the x-, y- and z-axes of the simulation box. In order to evaluate the effect of grid-locked galaxies on the galaxy–filament alignment signal, we have removed galaxies whose spin is comprised within less than 10° of any of the Cartesian planes of the box. Fig. B3 shows that the alignment signal without grid-locked galaxies is comparable to the case where all galaxies are accounted for. Low-mass galaxies have spin

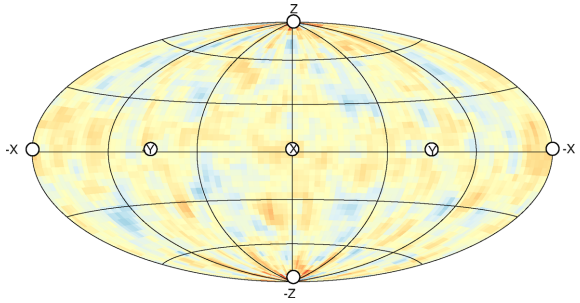


Figure B2. The effect of grid-locking of the skeleton segments at $z = 1.3$. The colour coding is the same as in Fig. B1. No preferred direction is shown for the skeleton.

preferentially aligned with their filament, and massive galaxies have a spin perpendicular to the filament with a transition mass between $10^{10.25}$ and $10^{10.75}$ $M_{\odot}$. This behaviour is expected as filaments do not suffer from grid-locking; the effect of grid-locking on low-mass galaxies brings some extra noise to the alignment measurement. Thus, the signal obtained for alignment of low-mass galaxies, while probably underestimated, is a robust trend. The same is true for high-mass galaxies that do not suffer from spurious grid-locking.

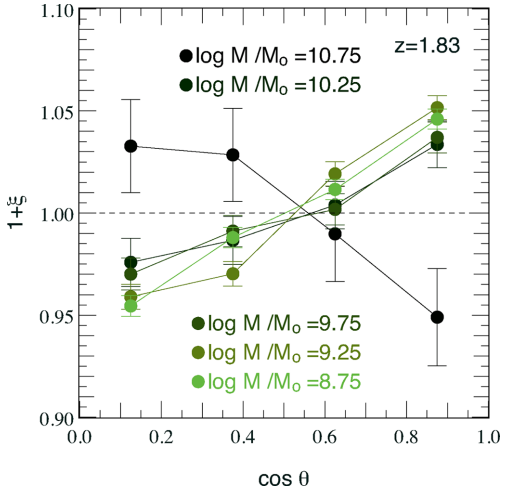


Figure B3. Excess probability ξ of the alignment between the spin of galaxies and their filament for galaxies with different stellar masses. Galaxies whose spin is contained within an angle smaller than 10° from any Cartesian planes of the box are not taken into account. Dashed line is the zero excess probability $\xi = 0$.

This paper has been typeset from a TEX/LATEX file prepared by the author.

PHYSICAL REVIEW D **86,** 103528 (2012)

Direct and fast calculation of regularized cosmological power spectrum at two-loop order

Atsushi Taruya,[1,2] Francis Bernardeau,[3] Takahiro Nishimichi,[2] and Sandrine Codis[4]

[1]*Research Center for the Early Universe, School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*
[2]*Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), Todai Institutes for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[3]*Institut de Physique Théorique, CEA, IPhT, F-91191 Gif-sur-Yvette, France and CNRS, URA 2306, F-91191 Gif-sur-Yvette, France*
[4]*Institut d'Astrophysique de Paris, 98 bis boulevard Arago, 75014 Paris, France*

(Received 15 August 2012; published 26 November 2012)

We present a specific prescription for the calculation of cosmological power spectra, exploited here at two-loop order in perturbation theory, based on the multipoint propagator expansion. In this approach, density and velocity power spectra are constructed from the regularized expressions of the propagators that reproduce both the resummed behavior in the high-k limit and the standard perturbation theory results at low k. With the help of N-body simulations, we particularly focus on the density field, and show that such a construction gives robust and accurate predictions for both the density power spectrum and the correlation function at percent level in the weakly nonlinear regime. We then present an algorithm that allows accelerated evaluations of all the required diagrams by reducing the computational tasks to one-dimensional integrals. This is achieved by means of precomputed kernel sets defined for appropriately chosen fiducial models. The computational time for two-loop results is then reduced from a few minutes, with the direct method, to a few seconds with the fast one. The robustness and applicability of this method are tested against the power spectrum COSMIC EMULATOR from which a wide variety of cosmological models can be explored. The FORTRAN PROGRAM with which direct and fast calculations of density power spectra can be done, REGPT, is publicly released as part of this paper.

DOI: 10.1103/PhysRevD.86.103528 PACS numbers: 98.80.−k, 98.65.Dx

I. INTRODUCTION

Since recombination, the large-scale structure of the Universe has evolved dominantly under the influence of both the cosmic expansion and the force of gravity acting on a pressureless fluid. The statistical nature of its spatial clustering is therefore expected to bring valuable cosmological information about the dynamics of the cosmic expansion and structure formation. Of particular importance is the measurement of baryon acoustic oscillations (BAOs) imprinted on the power spectrum or two-point correlation function (e.g., Refs. [1–5]) from which one can precisely determine the cosmological distance to the high-redshift universe, and henceforth clarify the nature of late-time cosmic acceleration (e.g., Refs. [6–10]). Precious information regarding the growth of structure are and will also be obtained from redshift-space distortions (e.g., Refs. [11–15]) and weak lensing measurements (see Refs. [16,17] and review papers [18,19]) at scales ranging to the linear or quasilinear to the nonlinear regimes. This could be captured with unprecedented details with the ongoing and future surveys, thanks to their redshift depth and large angular area, such as the Sloan Digital Sky Survey III,[1] the WiggleZ survey,[2] the Subaru Measurement of Imaging and Redshifts,[3] the Dark Energy Survey,[4] the BigBOSS project,[5] the Physics of the Accelerating Universe collaboration[6] and the ESA/Euclid survey.[7]

With the advent of such wealth of observations, there is therefore a growing interest in the development of theoretical tools to accurately compute the statistical quantities of the large-scale structure. At decreasing redshift and scale, the evolution of the large-scale structure however deviates significantly from the linear theory prediction and nonlinear gravitational clustering effects have to be taken into account. While N-body simulations can be relied upon in specific cases, because of the range of scales to be covered and the variety of models to explore, they should be complemented by analytical investigations that aim at computing the statistical properties of the large-scale structure from first principles, henceforth extending the validity range of linear calculations. It is to be noted that even at the scale of BAOs, linear calculations and one-loop corrections from the standard perturbation theory (PT) perform poorly (see e.g., Refs. [20–22]) asking for more advanced PT calculations. A crucial remark is that while higher-order PT corrections need to be included to improve the performance of predictions, the applicable range of PT prediction largely depends on the PT scheme itself. Indeed, the standard PT treatment is known to have bad convergence properties, and it produces ill-behaved higher-order

[1]www.sdss3.org.
[2]wigglez.swin.edu.au.
[3]sumire.ipmu.jp/en/.
[4]www.darkenergysurvey.org.
[5]bigboss.lbl.gov/index.html.
[6]www.pausurvey.org.
[7]www.euclid-ec.org.

corrections. The improvement of perturbation theory is thus a critical issue for the scientific exploitation of the coming surveys. Various resummation schemes have been proposed in Refs. [20,22–33] that aim at improving upon standard schemes. The aim of this paper is not to compare them but to propose, and test, a specific scheme that can be used routinely in practice.

In this paper, we are particularly interested in one of the resummation treatments, advocated in Ref. [34]. In this approach, the standard PT expansion is reorganized by introducing the multipoint propagators. These are the ensemble average of the infinitesimal variation of the cosmic fields with respect to the initial conditions. A key property shown in the previous reference is that all the statistical quantities such as power spectra and bispectra can be reconstructed by an expansion series written solely in terms of the multipoint propagators. This is referred to as the multipoint propagator expansion or Γ expansion. The advantage of this approach is that the nonperturbative properties, which can be obtained in standard PT by summing up infinite series of PT expansions, are wholly encapsulated in the multipoint propagators, including the effect of vertex renormalization. Furthermore, the Γ expansion has been found to be valid not only for Gaussian initial conditions, but also for non-Gaussian ones [35]. The construction of accurate calculation scheme for power spectra and bispectra can then be split in pieces that can be tested separately.

The second key property that leads us to consider such objects is that their global shape, e.g., their whole k-dependence, can be computed in a perturbation theory context and compared to N-body results thanks to the high-k exponential damping tail they all exhibit [34,36]. All these properties make the multipoint propagators the most important building blocks in the Γ expansion and the focus of our modeling efforts. In the following we will in particular make full use of the novel regularization scheme proposed in Ref. [37] that allows to consistently interpolate between standard PT results at low k and the expected resummed behavior at high k. This scheme has been explicitly tested for the two-point propagators up to two-loop order in Ref. [38] and for (specific shapes of) the three-point propagators in Ref. [37].

The first objective of this paper is to present an explicit calculation of the nonlinear power spectrum and correlation function of the cosmic density field based on this regularized treatment. Of particular interest is the extent to which the proposed scheme for Γ expansion works beyond standard PT when corrections at next-to-next-to-leading, i.e., two-loop, order are included. Results will be checked with N-body simulations. We will see that the Γ expansion with the regularized treatment of propagators, which we hereafter call REGPT, has good convergence properties and agrees remarkably well with simulations in the weakly nonlinear regime. Though the applicable range of PT treatment is still restricted to a certain wavenumber on large scales, the present REGPT treatment including the two-loop order is found to entirely cover the scales of BAOs at any redshift.

The second objective of this paper is to design and exploit a method to accelerate the power spectrum computations. Power spectra calculations in the context of REGPT calculations are rather involved requiring multidimensional integrations that have to be done with time-consuming Monte Carlo calculations. Typically, computing the power spectrum at percent level from our scheme takes several minutes. While this is acceptable when a handful of models have to be computed, this is an obstacle when a large domain of parameter space has to be systematically explored. Making use of the Γ expansion functional form, we found though that it is possible to exploit a novel technique for accelerated calculation, in which only one-dimensional integrals need to be evaluated while ensuring the same precision as rigorous REGPT calculations. The bottom line of this approach is to see the resulting nonlinear power spectrum as a functional of the linear power spectrum and then Taylor-expand this form with respect to the linear spectrum shape. We found that for well-chosen fiducial models, it is sufficient to Taylor-expand to first order only. We are then led to prepare in advance a set of kernel functions encoding the REGPT results for well chosen fiducial models, whose normalizations are left floating, from which the REGPT predictions for the target model can be calculated. We publicly release the Fortran code, REGPT, as a part of this paper.[8]

The organization of this paper is as follows. We begin by recalling the basic equations for cosmic fluid and perturbation theory in Sec. II. We introduce the multipoint propagator and give the power spectrum expression based on the Γ expansion. With the regularized treatment of multipoint propagators, in Sec. III, we examine the the power spectrum calculations including the corrections up to the two-loop order, and investigate their UV and IR sensitivity in evaluating the PT kernels. Then, in Sec. IV, a detailed comparison between PT calculation and N-body simulation is presented, and the accuracy and range of validity of PT calculation is checked. Based on this, Sec. V describes in detail the method to accelerate the power spectrum calculations. Robustness and applicability of the accelerated REGPT calculations to a wide range of cosmological models are tested against power spectrum COSMIC EMULATOR code in Sec. VI. Finally, in Sec. VII, we conclude and explore practical extensions of this work. The description of the publicly released code, REGPT, is presented in Appendix B.

[8]The code is available at www-utap.phys.s.u-tokyo.ac.jp/~ataruya/regpt_code.html.

II. EQUATIONS OF MOTION AND THE Γ EXPANSION

A. Equations of motion

In what follows, we consider the evolution of cold dark matter (CDM) plus baryon systems neglecting the tiny fraction of (massive) neutrinos. Owing to the single-stream approximation of the collisionless Boltzmann equation, which is thought to be quite accurate an approximation on large scales, the evolution of the CDM plus baryon system can be treated as an irrotational and pressureless fluid system whose governing equations are continuity and Euler equations in addition to the Poisson equation (see Ref. [39] for review). In the Fourier representation, these equations are further reduced to a more compact form. Let us introduce the two-component multiplet (e.g., Ref. [20])

$$\Psi_a(\boldsymbol{k};t) = \left(\delta(\boldsymbol{k};t), -\frac{\theta(\boldsymbol{k};t)}{f(t)}\right), \tag{1}$$

where the subscript $a = 1, 2$ selects the density and the velocity components of CDM plus baryons, with δ and $\theta(\boldsymbol{x}) \equiv \nabla\cdot\boldsymbol{v}(\boldsymbol{x})/(aH)$, where a and H are the scale factor of the Universe and the Hubble parameter, respectively. The function $f(t)$ is given by $f(t) \equiv d\ln D(t)/d\ln a$, and the quantity $D(t)$ is the linear growth factor. Then, in terms of the new time variable $\eta \equiv \ln D(t)$, the evolution equation for the vector quantity $\Psi_a(\boldsymbol{k};t)$ becomes

$$\left[\delta_{ab}\frac{\partial}{\partial\eta} + \Omega_{ab}(\eta)\right]\Psi_b(\boldsymbol{k};\eta) = \int\frac{d^3\boldsymbol{k}_1 d^3\boldsymbol{k}_2}{(2\pi)^3}\delta_D(\boldsymbol{k}-\boldsymbol{k}_1-\boldsymbol{k}_2)\gamma_{abc}(\boldsymbol{k}_1,\boldsymbol{k}_2) \times \Psi_b(\boldsymbol{k}_1;\eta)\Psi_c(\boldsymbol{k}_2;\eta), \tag{2}$$

where we used the summation convention, that is the repetition of the same subscripts indicates the sum over the whole multiplet components. In the above, the quantity δ_D is the Dirac delta function, and the time-dependent matrix $\Omega_{ab}(\eta)$ is given by

$$\Omega_{ab}(\eta) = \begin{pmatrix} 0 & -1 \\ -\frac{3}{2f^2}\Omega_{\rm m}(\eta) & \frac{3}{2f^2}\Omega_{\rm m}(\eta) - 1 \end{pmatrix} \tag{3}$$

with the quantity $\Omega_{\rm m}(\eta)$ being the density parameter of CDM plus baryons at a given time. The vertex function γ_{abc} becomes

$$\gamma_{abc}(\boldsymbol{k}_1,\boldsymbol{k}_2) = \begin{cases} \frac{1}{2}\left\{1 + \frac{\boldsymbol{k}_2\cdot\boldsymbol{k}_1}{|\boldsymbol{k}_2|^2}\right\}; & (a,b,c) = (1,1,2) \\ \frac{1}{2}\left\{1 + \frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{|\boldsymbol{k}_1|^2}\right\}; & (a,b,c) = (1,2,1) \\ \frac{(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)|\boldsymbol{k}_1+\boldsymbol{k}_2|^2}{2|\boldsymbol{k}_1|^2|\boldsymbol{k}_2|^2}; & (a,b,c) = (2,2,2) \\ 0; & \text{otherwise.} \end{cases} \tag{4}$$

Equation (2) can be recast as the integral equation (e.g., Refs. [20,39])

$$\Psi_a(\boldsymbol{k};\eta) = g_{ab}(\eta,\eta_0)\phi_b(\boldsymbol{k}) + \int_{\eta_0}^{\eta} d\eta' g_{ab}(\eta,\eta')\int\frac{d^3\boldsymbol{k}_1 d^3\boldsymbol{k}_2}{(2\pi)^3} \times \delta_D(\boldsymbol{k}-\boldsymbol{k}_1-\boldsymbol{k}_2)\gamma_{bcd}(\boldsymbol{k}_1,\boldsymbol{k}_2)\Psi_c(\boldsymbol{k}_1;\eta') \times \Psi_d(\boldsymbol{k}_2;\eta'). \tag{5}$$

The quantity $\phi_a(\boldsymbol{k}) \equiv \Psi_a(\boldsymbol{k},\eta_0)$ denotes the initial condition, and the function g_{ab} denotes the linear propagator satisfying the following equation,

$$\left[\delta_{ab}\frac{\partial}{\partial\eta} + \Omega_{ab}(\eta)\right]g_{bc}(\eta,\eta') = 0, \tag{6}$$

with the boundary condition $g_{ab}(\eta,\eta) = \delta_{ab}$. The statistical properties of the field Ψ_a are encoded in the initial field ϕ_a, for which we assume Gaussian statistics. The power spectrum of ϕ_a is defined as

$$\langle\phi_a(\boldsymbol{k})\phi_b(\boldsymbol{k}')\rangle = (2\pi)^3\delta_D(\boldsymbol{k}+\boldsymbol{k}')P_{ab,0}(k). \tag{7}$$

In what follows, most of the calculations will be made assuming the contribution of decaying modes of linear perturbation can be neglected. This implies that the field $\phi_a(\boldsymbol{k})$ is factorized as $\phi_a(\boldsymbol{k}) = \delta_0(\boldsymbol{k})u_a$ with $u_a = (1,1)$, and thus the initial power spectrum is written as $P_{ab,0}(k) = P_0(k)u_a u_b$.

Using the formal expression (5), a perturbative solution is obtained by expanding the fields in terms of the initial fields

$$\Psi_a(\boldsymbol{k};\eta) = \sum_{n=1}^{\infty}\Psi_a^{(n)}(\boldsymbol{k};\eta). \tag{8}$$

The expression of the solution at each order is written as

$$\Psi_a^{(n)}(\boldsymbol{k};\eta) = \int\frac{d^3\boldsymbol{k}_1\cdots d^3\boldsymbol{k}_n}{(2\pi)^{3(n-1)}}\delta_D(\boldsymbol{k}-\boldsymbol{k}_1-\ldots-\boldsymbol{k}_n) \times \mathcal{F}_a^{(n)}(\boldsymbol{k}_1,\boldsymbol{k}_2,\cdots,\boldsymbol{k}_n;\eta)\delta_0(\boldsymbol{k}_1)\cdots\delta_0(\boldsymbol{k}_n). \tag{9}$$

The kernel $\mathcal{F}_a^{(n)}$ is generally a complicated time-dependent function, but can be constructed in terms of the quantities γ_{abc} and g_{ab}. Examples of the solutions are shown diagrammatically in Fig. 1. Because we are interested in the

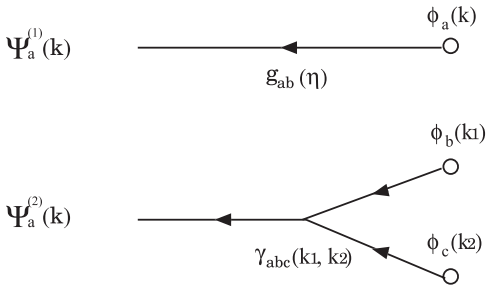


FIG. 1. Diagrammatic representation of the standard PT expansion.

late-time evolution of large-scale structure only, we can take the limit $\eta_0 \to -\infty$. As a consequence, the fastest growing term is the only surviving one and the kernel is simplified into

$$\mathcal{F}_a^{(n)}(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_n) = e^{n\eta} F_{a,\mathrm{sym}}^{(n)}(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_n), \tag{10}$$

where the function $F_{a,\mathrm{sym}}^{(n)}$ is the symmetrized standard PT kernel, sometimes written as $F_{a,\mathrm{sym}}^{(n)} = (F_n, G_n)$, whose explicit expressions are obtained from recursion relations as recalled in Ref. [39].

B. Γ expansion and regularized PT treatment

In this paper, we are more specifically interested in the power spectra $P_{ab}(k;\eta)$, defined as

$$\langle \Psi_a(\boldsymbol{k};\eta)\Psi_b(\boldsymbol{k}';\eta)\rangle = (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}+\boldsymbol{k}')P_{ab}(|\boldsymbol{k}|;\eta). \tag{11}$$

Substituting a set of perturbative solutions (9) into the above definition, it is straightforward to obtain the successive perturbative expressions for the power spectra. This is the standard PT treatment where the initial fields values are seen as the perturbative variables. The standard PT calculations have, however, been shown to produce ill-behaved higher-order corrections that lack good convergence properties.

As an alternative to the standard PT framework, it has been recently advocated by many authors that the PT expansion can be reorganized by introducing nonperturbative quantities to improve the resulting convergence of the expansion. The Γ expansion is one such nonperturbative framework, and the so-called multipoint propagators constitute the building blocks of this Γ expansion. Denoting the $(p+1)$-point propagator by $\Gamma^{(p)}$, we define

$$\frac{1}{p!}\left\langle \frac{\delta^p \Psi_a(\boldsymbol{k},\eta)}{\delta\phi_{c_1}(\boldsymbol{k}_1)\cdots\delta\phi_{c_p}(\boldsymbol{k}_p)}\right\rangle = \delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}_{1\ldots p})\frac{1}{(2\pi)^{3(p-1)}}\Gamma_{ac_1\cdots c_p}^{(p)}(\boldsymbol{k}_1,\cdots,\boldsymbol{k}_p;\eta). \tag{12}$$

With these objects, the power spectra is shown to be expressed as [34],

$$P_{ab}(|\boldsymbol{k}|;\eta) = \sum_{t=1}^{\infty} t! \int \frac{d^3\boldsymbol{q}_1\cdots d^3\boldsymbol{q}_t}{(2\pi)^{3(t-1)}}\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{q}_{1\ldots t}) \times \Gamma_a^{(t)}(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_t;\eta)\Gamma_b^{(t)}(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_t;\eta) \times P_0(q_1)\cdots P_0(q_t), \tag{13}$$

where we introduced the shorthand notation,

$$\Gamma_a^{(t)}(\boldsymbol{q}_1,\ldots,\boldsymbol{q}_t;\eta) = \Gamma_{ac_1\cdots c_t}^{(t)}(\boldsymbol{q}_1,\ldots,\boldsymbol{q}_t;\eta)u_{c_1}\cdots u_{c_t}. \tag{14}$$

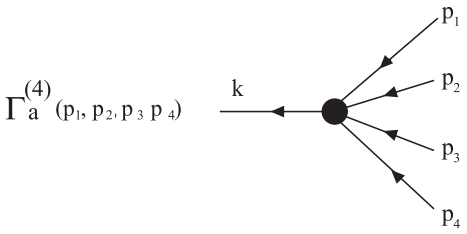


FIG. 2. Example of the multipoint propagator, $\Gamma_a^{(4)}$. A large filled circle symbolically represents all possible contributions that enter into the fully nonlinear propagator. A part of those contributions can be seen graphically using PT expansion (see Figs. 4 and 5 for three-point propagator $\Gamma_a^{(2)}$).

The diagrammatic representation for multipoint propagator and the power spectrum is respectively shown in Figs. 2 and 3.

The construction of the Γ expansion is rather transparent, and like Eq. (13), one easily finds the expressions for the higher-order statistical quantities such as bispectrum. Another important point is that one can exploit the asymptotic properties of the propagators $\Gamma^{(p)}$ beyond perturbation theory expansions. To be precise, in the high-k limit, higher-order contributions can be systematically computed at all orders, and as a result of summing up all the contributions, the multipoint propagators are shown to be exponentially suppressed [34,36]

$$\Gamma_a^{(p)}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta) \overset{k\to\infty}{\to} \exp\left\{-\frac{k^2\sigma_{\mathrm{d}}^2 e^{2\eta}}{2}\right\}\Gamma_{a,\mathrm{tree}}^{(p)}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta) \tag{15}$$

with $k = |\boldsymbol{k}_1 + \cdots + \boldsymbol{k}_p|$. This is the generalization of the result for the two-point propagator in Ref. [23]. Here, the quantity $\Gamma_{a,\mathrm{tree}}^{(p)}$ is the lowest-order nonvanishing propagator obtained from the standard PT calculation, and σ_{d} is the one-dimensional root mean square of the displacement field defined by

$$\sigma_{\mathrm{d}}^2 = \frac{1}{3}\int \frac{d^3\boldsymbol{q}}{(2\pi)^3}\frac{P_0(q)}{q^2}. \tag{16}$$

The form of Eq. (15) does not however provide a good description of the propagators at all scale. At low k the propagators are expected to approach their standard PT expressions that can be written formally,

$$\Gamma_a^{(p)}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta) = \Gamma_{a,\mathrm{tree}}^{(p)}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta) + \sum_{n=1}^{\infty}\Gamma_{a,n-\mathrm{loop}}^{(p)}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta). \tag{17}$$

For the dominant growing-mode contribution we are interested in, each correction term is expressed in terms of the standard PT kernels as,

$$\Gamma_{a,\mathrm{tree}}^{(p)}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_2;\eta) = e^{p\eta}F_{a,\mathrm{sym}}^{(p)}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p), \tag{18}$$

for the tree-level contribution, and

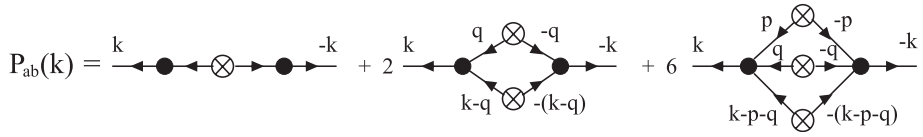


FIG. 3. Diagrammatic representation of the power spectrum by means of Γ expansion. Here, the result up to the two-loop order is shown. In each contribution of the diagrams, the multipoint propagators are *glued* together at the crossed circles where the initial power spectra $P_0(k)$ are inserted.

$$\Gamma^{(p)}_{a,n-\mathrm{loop}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta) = e^{(2n+p)\eta} c_n^{(p)} \int \frac{d^3\boldsymbol{p}_1\cdots d^3\boldsymbol{p}_n}{(2\pi)^{3n}} F^{(2n+p)}_{a,\mathrm{sym}}(\boldsymbol{p}_1,-\boldsymbol{p}_1,\cdots,\boldsymbol{p}_n,-\boldsymbol{p}_n,\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p) P_0(p_1)\cdots P_0(p_n)$$
$$\equiv e^{(2n+p)\eta}\bar{\Gamma}^{(p)}_{a,n-\ \mathrm{loop}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p) \tag{19}$$

for the n-loop order contributions, where the coefficient $c_n^{(p)}$ is given by $c_n^{(p)} = {}_{(2n+p)}C_p(2n-1)!!$ with ${}_{(2n+p)}C_p$ being the binomial coefficient. The graphical representation of the standard PT expansion is shown in Fig. 4. The important remark in Eq. (19) is that each perturbative correction possesses the following asymptotic form:

$$\Gamma^{(p)}_{a,n-\mathrm{loop}} \overset{k\to\infty}{\longrightarrow} \frac{1}{n!}\left(-\frac{k^2\sigma_{\mathrm{d}}^2 e^{2\eta}}{2}\right)^n \Gamma^{(p)}_{a,n-\mathrm{tree}}, \tag{20}$$

which consistently recovers the expression (15) when we sum up all the loop contributions. This indicates the existence of a matching scheme which smoothly interpolates between the low-k and high-k results for any multipoint propagator. Such a scheme has been proposed in Ref [37] where a novel regularized scheme, in which the low- and high-k behaviors are jointly reproduced, is derived. The construction of the *regularized* propagator is totally unambiguous. They can incorporate an arbitrary number of loop corrections.

Restricting the results to the growing mode contributions, the *regularized* propagators are expressed in a transparent way in terms of the standard PT results, and one gets

$$\Gamma^{(p)}_{a,\mathrm{reg}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta)$$
$$= e^{p\eta}\left[F^{(p)}_{a,\mathrm{sym}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p)\left\{1+\frac{k^2\sigma_{\mathrm{d}}^2 e^{2\eta}}{2}\right\} + e^{2\eta}\bar{\Gamma}^{(p)}_{a,1-\mathrm{loop}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p)\right]\exp\left\{-\frac{k^2\sigma_{\mathrm{d}}^2 e^{2\eta}}{2}\right\}, \tag{21}$$

which consistently reproduces one-loop PT results at low k. An example of the regularized propagator valid at one-loop order is diagrammatically shown in Fig. 5. This

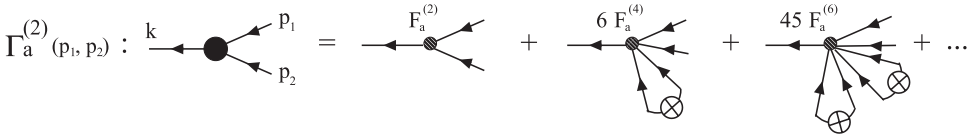


FIG. 4. Diagrammatic representation of the standard PT expansion for three-point propagator, $\Gamma^{(2)}_a$. For fastest growing-mode contribution, the standard PT kernels, $F^{(n)}_{a,\mathrm{sym}}$, form the basic pieces of PT expansion, depicted as incoming lines connected to a single outgoing line at the shaded circle. In the case of $\Gamma^{(2)}_a$, the leading-order contribution is $F^{(2)}_{a,\mathrm{sym}}$, and successively the kernels $F^{(4)}_{a,\mathrm{sym}}$ and $F^{(6)}_{a,\mathrm{sym}}$ appear as higher-order contributions, for which pairs of the incoming lines are glued at the crossed circle, which indicates the initial power spectrum P_0, forming closed loops.

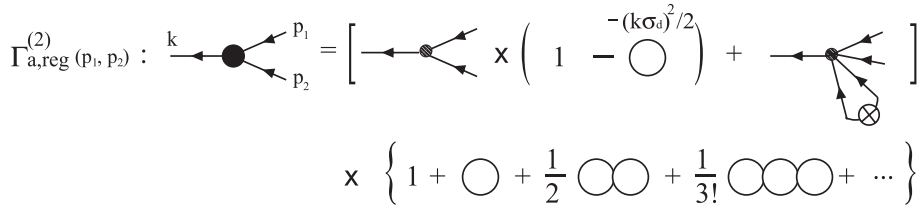


FIG. 5. Diagrammatic representation of the regularized three-point propagator, $\Gamma^{(2)}_{a,\mathrm{reg}}$. In the high-$k$ limit, the higher-loop contribution for three-point propagator behaves like Eq. (20), indicating that each loop diagram ($\Gamma^{(2)}_{a,n-\ \mathrm{loop}}$) is effectively split into tree diagram ($F^{(2)}_{a,\mathrm{sym}}$) and self-loop diagram ($\{-(k\sigma_{\mathrm{d}})^2/2\}^n/n!$), the latter of which is depicted as open loops. Systematically summing up all the higher-loop contributions, we recover Eq. (15), which is graphically shown as the infinite sum of open-loop diagrams in the brace. To reproduce the standard PT result at low k, the one-loop diagram is inserted in the bracket, and the tree diagram is multiplied by the counter term $\{1+(k\sigma_{\mathrm{d}})^2/2\}$.

construction is easily generalized to include the higher-order PT corrections at low k. For instance, the regularized propagator including the corrections up to the two-loop order becomes

$$\Gamma^{(p)}_{a,\mathrm{reg}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p;\eta) = e^{p,\eta}\Big[F^{(p)}_{a,\mathrm{tree}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p)\Big\{1+\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}+\frac{1}{2}\Big(\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big)^2\Big\} + e^{2\eta}\bar{\Gamma}^{(p)}_{1-\mathrm{loop}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p)\Big\{1+\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big\} + e^{4\eta}\bar{\Gamma}^{(p)}_{2-\mathrm{loop}}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p)\Big]\exp\Big\{-\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big\}. \quad (22)$$

Note that the functions $\bar{\Gamma}^{(p)}_{n\text{-loop}}$ are the scale-dependent part of the propagator defined by Eq. (19).

III. POWER SPECTRUM CALCULATION FROM REGULARIZED Γ EXPANSION

A. Power spectrum at two-loop order

Since the proposed regularized propagators preserve the expected low-k and high-k behaviors, the convergence of the Γ expansion adopting the regularization scheme would be much better than the standard PT expansion. In this paper, applying this regularized PT treatment, we will explicitly demonstrate the power spectrum calculations at two-loop order. Comparing those predictions with N-body simulations, the validity and precision of PT treatment are discussed. From Eq. (13), the explicit expression for the power spectrum valid up to the two-loop order becomes

$$P_{ab}(k;\eta) = \Gamma^{(1)}_{a,\mathrm{reg}}(k;\eta)\Gamma^{(1)}_{b,\mathrm{reg}}(k;\eta)P_0(k) + 2\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}\Gamma^{(2)}_{a,\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)\Gamma^{(2)}_{b,\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)P_0(q)P_0(|\boldsymbol{k}-\boldsymbol{q}|) + 6\int\frac{d^6\boldsymbol{p}d^3\boldsymbol{q}}{(2\pi)^6}\Gamma^{(3)}_{a,\mathrm{reg}}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q};\eta)\Gamma^{(3)}_{b,\mathrm{reg}}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q};\eta)P_0(p)P_0(q)P_0(|\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}|) \quad (23)$$

with the regularized propagators given by

$$\Gamma^{(1)}_{a,\mathrm{reg}}(k;\eta) = e^{\eta}\Big[1+\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}+\frac{1}{2}\Big(\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big)^2 + e^{2\eta}\bar{\Gamma}^{(1)}_{a,1-\mathrm{loop}}(k)\Big\{1+\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big\} + e^{4\eta}\bar{\Gamma}^{(1)}_{a,2-\mathrm{loop}}(k)\Big]\exp\Big\{-\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big\}, \quad (24)$$

$$\Gamma^{(2)}_{a,\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta) = e^{2\eta}\Big[F^{(2)}_{a,\mathrm{sym}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})\Big\{1+\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big\} + e^{2\eta}\bar{\Gamma}^{(2)}_{a,1-\mathrm{loop}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})\Big] \times\exp\Big\{-\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big\}, \quad (25)$$

$$\Gamma^{(3)}_{a,\mathrm{reg}}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q};\eta) = e^{3\eta}F^{(3)}_{a,\mathrm{sym}}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}) \times\exp\Big\{-\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\Big\}. \quad (26)$$

Note that the higher-order contributions up to the two- and one-loop order of the propagators are respectively included in the expression of the regularized propagators $\Gamma^{(1)}_{a,\mathrm{reg}}$ and $\Gamma^{(2)}_{a,\mathrm{reg}}$, consistently with the Γ expansion at two-loop order.

The power spectrum expression involves many integrals, but, most of them are reduced to two- or three-dimensional integrals if one uses the analytic expressions for the kernels of higher-loop corrections $\bar{\Gamma}$ in the regularized propagator. We use the expression in Ref. [38] to evaluate $\bar{\Gamma}^{(2)}_{1\text{-loop}}$, and adopt the fitting functions for the kernel of $\bar{\Gamma}^{(1)}_{2\text{-loop}}$ (see Ref. [38]). We then apply the method of Gaussian quadrature to the numerical evaluation of the low-dimensional integrals. A bit cumbersome is the integral containing $\Gamma^{(3)}$. While it can be reduced to a four-dimensional integral in principle, the expression of the resulting kernel would be very cumbersome and might not be suited for practical calculation. We thus adopt the Monte Carlo technique of quasirandom sampling using the CUBA library [40], and evaluate the five-dimensional integral directly.[9]

Fig. 6 illustrates an example how each correction term in the regularized Γ expansion contributes to the total power spectrum. The plotted result is the density power spectrum, P_{11}, and the contribution of the term involving each multi-point propagator is separately shown. The three corrections contribute to the power spectrum at different scales, and the higher-order terms involving $\Gamma^{(2)}_{\mathrm{reg}}$ and $\Gamma^{(3)}_{\mathrm{reg}}$ are well localized, each producing one bump. This is a clear manifestation of the result of the regularized PT treatment, and it resembles what the renormalized perturbation theory calculations by Ref. [24] give.

In the next section, the results of the regularized Γ expansion will be compared with N-body simulations. But, before doing that, we will give several remarks and comments on the computation of the power spectrum in the subsequent subsection.

[9]Since the final result of the integration is expressed as a function of only the wavenumber k, the integrand possesses an azimuthal symmetry with respect to the vector, $\boldsymbol{k}$, indicating that the integral is reduced to a five-dimensional integral.

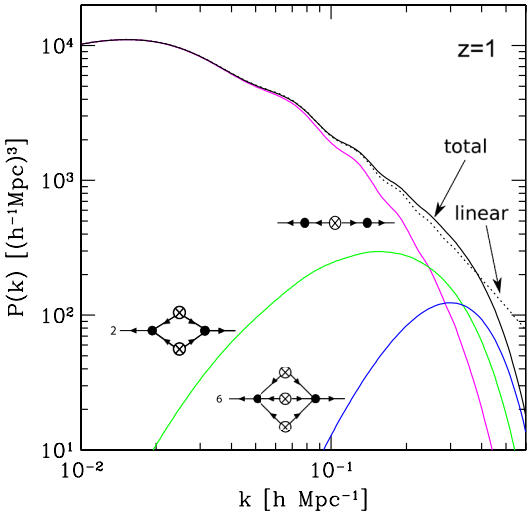


FIG. 6 (color online). Contribution of multipoint propagators to the power spectrum, $P(k) = P_{11}(k)$ at $z = 1$. Magenta, green, and blue curves represent the power spectrum contributions from the first, second, and third terms at the right-hand side of Eq. (23), respectively, each of which just corresponds to the diagram in Fig. 3, involving $\Gamma^{(1)}_{\rm reg}$, $\Gamma^{(2)}_{\rm reg}$, and $\Gamma^{(3)}_{\rm reg}$. Summing up these contributions, total power spectrum is shown in black solid line. For reference, linear power spectrum is also plotted as dotted line.

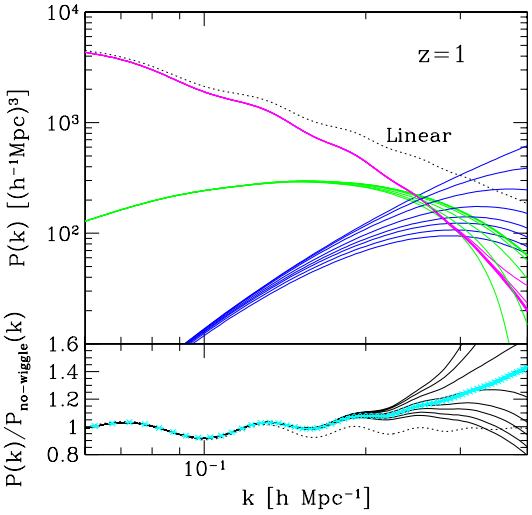


FIG. 7 (color online). Sensitivity of the power spectrum prediction at $z = 1$ to the UV cutoff in the estimation of $\sigma_{\rm d}$. Top panel shows each contribution of the power spectrum corrections involving $\Gamma^{(1)}$ (magenta), $\Gamma^{(2)}$ (green), and $\Gamma^{(3)}$ (blue), respectively (from top to bottom, see also Fig. 6). Bottom panel shows the total sum of power spectrum divided by the smooth reference power spectrum, $P_{\rm no\text{-}wiggle}(k)$, which is calculated from the no-wiggle formula of the linear transfer function in Ref. [47]. In each case, top lines represent the results obtained by setting $\sigma_{\rm d} = 0$, while undermost lines show the cases adopting the value of $\sigma_{\rm d}$ without UV cutoff. The middle six lines represent the cases adopting the running UV cutoff in estimating $\sigma_{\rm d}$, with cutoff $k_\Lambda(k) = k, k/2, k/3, k/5, k/10$, and $k/20$ (from bottom to top). As a reference, linear theory prediction is also plotted in both panels (dotted).

B. Effect of running UV cutoff for $\sigma_{\rm d}$

Since the shape of the power spectrum given by Eq. (23) significantly depends on the exponential damping in the regularized propagators, we first comment on the effect of this function. As it has been shown, the exponential function arises from the summation of infinite series of perturbations at all order in the high-k limit. Recently, Ref. [36] advocated that this exponential function can be interpreted as the result of resummation at hard part (high k), and the displacement dispersion $\sigma_{\rm d}$ in the exponent must be evaluated in a consistent way that the domain of the integral is restricted to a soft part (low k). This implies that depending on the scale of our interest, the boundary of the soft and hard domains can be changed, and the resulting quantity $\sigma_{\rm d}$ should be regarded as a scale-dependent function.

In Fig. 7, we examine the impact of the scale-dependent $\sigma_{\rm d}$ on the power spectrum at $z = 1$. Plotted results are the contributions of the power spectrum corrections (upper) and the total power spectrum divided by the smooth reference linear spectrum (bottom). Here, we evaluate $\sigma_{\rm d}$ by introducing the running UV cutoff $k_\Lambda(k)$

$$\sigma_{\rm d}^2(k) = \int_0^{k_\Lambda(k)} \frac{dq}{6\pi^2} P_0(q). \tag{27}$$

Various curves in Fig. 7 represent the results with different prescription for the running UV cutoff. The correction involving the four-point propagator $\Gamma^{(3)}_{\rm reg}$ is most sensitively affected by the running cutoff, and the resulting power spectrum significantly varies at scales $k \gtrsim 0.2h$ Mpc^{-1}. This is because the exponential damping manifests itself at the scale $k \sim 1/(e^\eta \sigma_{\rm d})$ where the contribution from $\Gamma^{(3)}_{\rm reg}$, which contains no relevant terms counteracted with the exponential damping, becomes significant among the three corrections.

In the bottom panel of Fig. 7, we also plot the result of N-body simulations (see Sec. IV A). The comparison with simulation suggests that the PT calculation with running cutoff $k_\Lambda \sim k/2 - k/5$ is favored, although there is no clear physical reason why this is so. Strictly speaking, the running IR cutoff might also be introduced in evaluating all the integrals in the power spectrum expression, so as to consistently discriminate between the contributions coming from soft and hard parts. Moreover, the running cutoff k_Λ may also depend on the redshift. These complications mostly come from the ambiguity of the boundary between soft and hard domains in our regularization scheme. For practical purpose to the cosmological application, we postpone these issues to future investigation, and take a rather phenomenological approach. Hereafter, the running cutoff is only introduced in evaluating $\sigma_{\rm d}$, and we evaluate it according to Eq. (27), setting the cutoff scale to $k_\Lambda = k/2$. With this treatment, we will see later that the PT prediction becomes improved compared to the standard

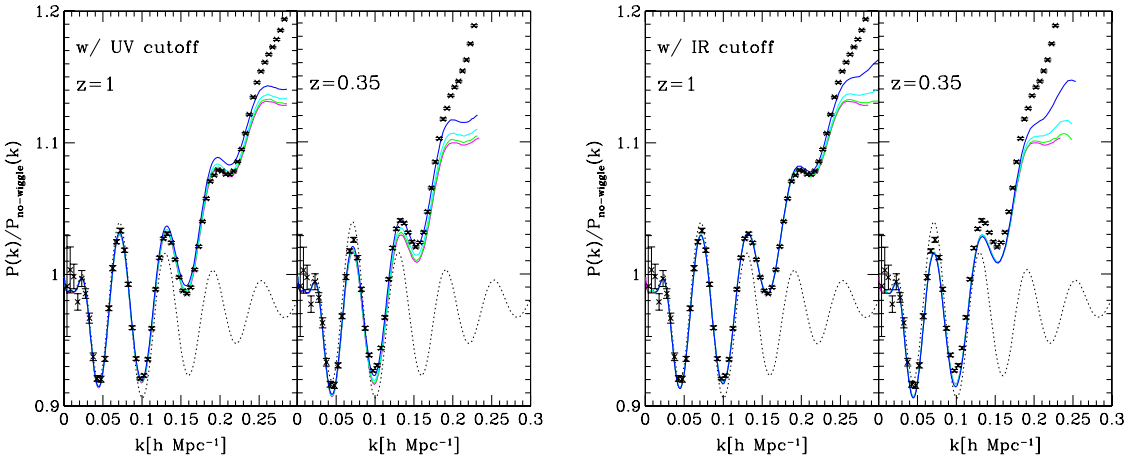


FIG. 8 (color online). Sensitivity of the power spectrum prediction to the UV (left) and IR (right) cutoff. In each panel, the ratio of power spectrum, $P(k)/P_{\text{no-wiggle}}(k)$, is plotted as function of k at $z = 1$ (left) and 0.35 (right). In evaluating the integrals of the power spectrum corrections, the maximum wavenumber for the range of the integral is set to 2π (green), π (cyan), and $\pi/2$ (blue) h Mpc^{-1} in the left panel, while in the right panel, we change the minimum wavenumber k_{min} to $k_{\text{min}} = 2\pi/L_{\text{box}}$ with $L_{\text{box}} = 2048$ (green), 1024 (cyan), and 512 h Mpc^{-1} (blue) (from second lowest to top). The magenta lines (bottom) indicate the results adopting the default set of parameters $(k_{\text{min}}, k_{\text{max}}) = (5 \times 10^{-4},\ 10)\ h$ Mpc^{-1}.

PT calculation, and it reproduces the N-body results quite well at any redshift.

C. Sensitivity to IR and UV cutoff

In computing the power spectrum, except for σ_{d}, the domains of each integral in Eq. (23) are usually taken broadly enough so as to ensure the convergence of the results. In comparison with N-body simulations, however, a care must be taken because the available Fourier modes in simulations are restricted depending on the simulation box size and/or mesh size of Fourier transform, which affect both the efficiency of mode transfer and strength of mode coupling. The evolved result of the power spectrum would thus be changed, and it should be carefully compared with PT calculation, taking the finite resolution into consideration. Here, focusing on the BAO scales, we briefly discuss the sensitivity of the PT calculation to the IR and UV cutoff in the integrals.

Figure 8 shows the variation of the power spectra with respect to the UV (left) and IR (right) cutoff. In general, the kernel of integrals becomes broader for higher-loop corrections, and thus the two-point propagator $\Gamma^{(1)}_{\text{reg}}$ containing the two-loop contribution is sensibly affected by the UV cutoff. Note that the signs of one- and two-loop corrections in $\Gamma^{(1)}_{\text{reg}}$ are opposite at BAO scales. Hence, as decreasing the cutoff wavenumber k_{max}, the cancellation of each contribution is relatively relaxed, and the power spectrum amplitude gets increased. On the other hand, due to the lack of the long-wave modes, the IR cutoff not only decreases each contribution of the loop integrals, but also reduces σ_{d}, leading to a slight suppression of the exponential damping. The net effect of the IR cutoff, especially at small scales $k \gtrsim 0.2h$ Mpc^{-1}, is that the latter overcomes the former, and the total power spectrum is slightly enhanced.

These results imply that the effect of UV and IR cutoff not only affects the power spectrum shape at small scales, but also causes a slight offset in power spectrum amplitude at moderately large scales, $k \sim 0.1h$ Mpc^{-1}. The size of these effects is basically small, but would not be negligible in a percent-level comparison. Based on this remark, in what follows, we adopt the cutoff scales $(k_{\text{min}}, k_{\text{max}}) = (5 \times 10^{-4}, 10)h$ Mpc^{-1} as default parameters to compute the power spectra. With this setup, REGPT calculation gives a mostly convergent result, which can be compared with high-resolution N-body simulations with a large box size.

D. Comparison with MPTbreeze

In Ref. [41], MPTBREEZE, an alternative scheme has been proposed for the construction of power spectra that is based on the same multipoint propagator expansion. This proposition is, however, based on simplified assumptions regarding the behavior of the multipoint propagators. More specifically, in MPTBREEZE, the propagators are assumed to take the form

$$\Gamma^{(p)}_{a,\text{reg}}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{p-1}, \boldsymbol{k} - \boldsymbol{q}_{1\cdots(p-1)}; \eta) = e^{p\eta} F^{(p)}_{a,\text{sym}}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{p-1}, \boldsymbol{k} - \boldsymbol{q}_{1\cdots(p-1)}) \exp\{f_a(k) e^{2\eta}\}, \tag{28}$$

where $f_1(k)$ and $f_2(k)$ are the one-loop corrections to the density and velocity propagators, respectively. This form corresponds to the late-time original expression of the exponentiation scheme initially put forward in Ref. [23]. It is shown in Refs. [37,38] that at one-loop order, this prescription gives nearly identical result for the two-point

propagator to the prescription proposed in Ref. [23]. The MPTBREEZE prescription, however, ignores the impact of two-loop PT corrections on the two-point propagators. From the results presented in Ref. [38], it implies that MPTBREEZE might be outperformed by REGPT at $z \gtrsim 1$. On the other hand, the predictions of that scheme are made more robust because they are less sensitive to the UV part of the linear spectrum as discussed in that paper. Furthermore, the one-loop correction for the three- and four-point propagators is treated in an effective way. These simplified assumptions allow a more rapid calculations of the set of diagrams. It takes just a few seconds to get the expected shape in this scheme. The computational time is, however, rather comparable to the fast implementation of REGPT which we will present in Sec. V.

IV. COMPARISON WITH *N*-BODY SIMULATIONS

We are now in position to present quantitative comparisons between REGPT calculations and N-body simulations. After briefly describing the N-body simulations in Sec. IV A, we show the results of power spectrum and two-point correlation function in Sec. IV B and IV C, respectively. Precision and validity of the PT predictions are discussed in detail.

A. *N*-body simulations

To compare REGPT calculations with N-body simulations, we ran a new set of N-body simulations, which will be presented in more detail with an extensive convergence study in Ref. [42]. This set of simulations can be regarded as an updated version of the one presented in Ref. [22] with much larger total volume and a more careful setup to achieve a smaller statistical and systematic error. The data were created by a public N-body code GADGET2 [43] with cubic boxes of side length $2,048h^{-1}$ Mpc, and 1024^3 particles. The cosmological parameters adopted in these N-body simulations are basically the same as in the previous one, and are determined by the five-year WMAP observations [44] (see Table I). The initial conditions were generated by a parallelized version of the 2LPT code [45], developed in Ref. [46]. After several tests given in Ref. [46], a lower initial redshift $z_{\rm init}$ turns out to give a more reliable estimate for the power spectrum at BAO scales, and we thus adopt the initial redshift $z_{\rm init} = 15$. With this setup, we have created 60 independent realizations and the data were stored at redshifts $z = 3$, 2, 1, and 0.35. The total volume at each output redshift is $515h^{-3}$ Gpc^3, which is statistically sufficient for a detailed comparison with PT calculations.

We measure both the matter power spectrum and the correlation function. For the power spectrum, we adopt the cloud-in-cells interpolation, and construct the Fourier transform of the density field assigned on the $1,024^3$ grids. As for the estimation of the two-point correlation function, we adopt the grid-based calculation using the fast Fourier transformation [22]. Similarly to the power spectrum analysis, we first compute the square of the density field on each grid point in Fourier space. Then, applying the inverse Fourier transformation, we take the average over separation vectors and realizations, and finally obtain the two-point correlation function. The implementation of this method, together with a convergence test, is presented in more detail in Ref. [22].

B. Power spectrum

Let us first present the power spectrum results. Left panel of Fig. 9 shows the ratio of the power spectra, $P(k)/P_{\text{no-wiggle}}(k)$, while the right panel plots the fractional difference between N-body simulations and PT calculations, defined by $[P_{\text{N-body}}(k) - P_{\rm PT}(k)]/P_{\text{no-wiggle}}(k)$ (where $P_{\text{no-wiggle}}(k)$ is calculated from the no-wiggle formula of the linear transfer function in Ref. [47]). Overall, the agreement between REGPT and N-body simulations is remarkable at low k, and a percent-level agreement is achieved up to a certain wavenumber. For a decreasing redshift, the nonlinearities develop and the applicable range of PT calculations inevitably becomes narrower, however, compared to the standard PT predictions, the REGPT result can reproduce the N-body trend over an even wider range. Indeed, the range of agreement with N-body simulations is rather comparable to other improved PTs including higher-order corrections, such as closure theory [22,29], and better than some of those predictions. For reference, we compute the power spectra from closure and Lagrangian resummation theory (LRT) [25,48] at two-loop order, and estimate the range of a percent-level agreement with N-body simulations, the results of which are respectively depicted as green and blue vertical arrows in right panel of Fig. 9. Note that at $z = 3$, the range of agreement for closure theory exceeds the plotted range, and is not shown here.

Although the REGPT treatment gives a very good performance comparable to or even better than other improved PTs, a closer look at Fig. 9 reveals several subpercent discrepancies.

(i) One is the low-k behavior at $z = 0.35$, which exhibits a small discrepancy with N-body simulation. Our investigations indicate that it is probably due to a

TABLE I. Cosmological parameters for N-body simulations (ΛCDM).

Name	$L_{\rm box}$	No. of particles	$z_{\rm ini}$	No. of runs	$\Omega_{\rm m}$	Ω_Λ	$\Omega_{\rm b}/\Omega_{\rm m}$	w	h	n_s	σ_8
WMAP5	$2,048h^{-1}$ Mpc	$1,024^3$	15	60	0.279	0.721	0.165	-1	0.701	0.96	0.815_9

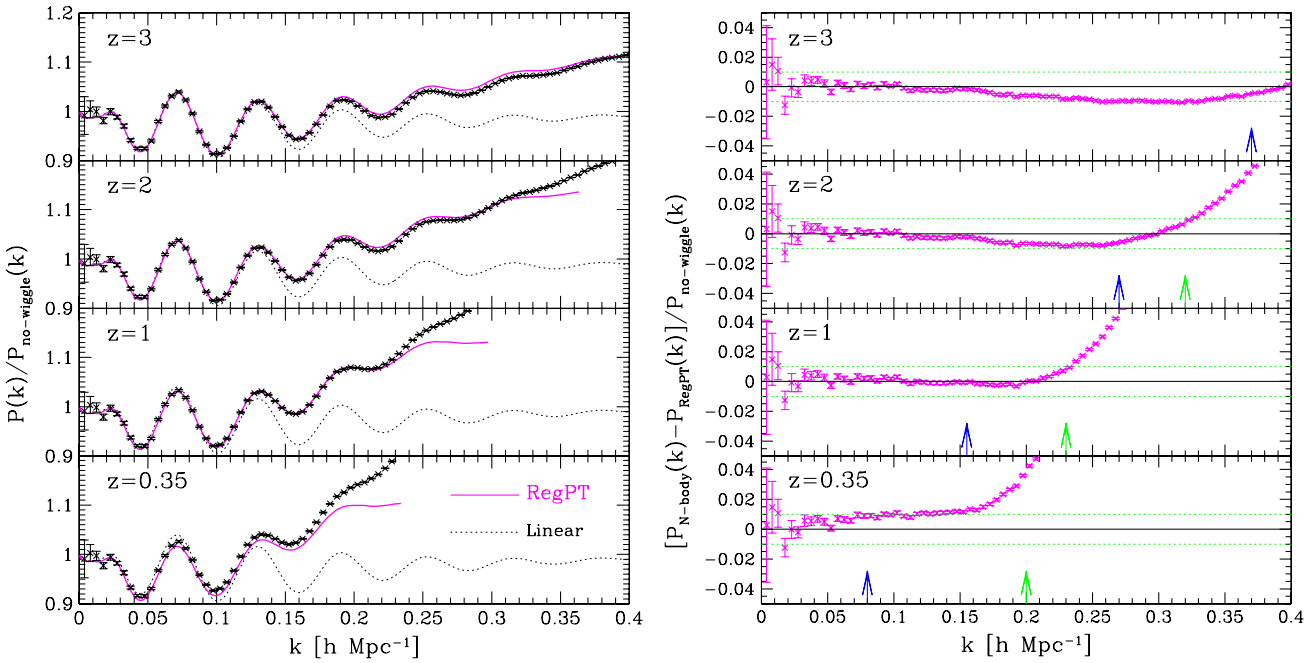


FIG. 9 (color online). Comparison of power spectrum results between N-body simulations and REGPT calculations. In each panel, the results at $z = 3$, 2, 1, and 0.35 are shown (from top to bottom). Left panel shows the ratio of power spectrum to the smooth linear spectrum, $P(k)/P_{\mathrm{no-wiggle}}(k)$, where the reference spectrum $P_{\mathrm{no-wiggle}}(k)$ is calculated from the no-wiggle formula of the linear transfer function in Ref. [47]. Solid lines are the REGPT results, while dotted lines represent the linear theory predictions. Right panel plots the difference between N-body and REGPT results normalized by the no-wiggle spectrum, i.e., $[P_{\mathrm{N-body}}(k) - P_{\mathrm{RegPT}}(k)]/P_{\mathrm{no-wiggle}}(k)$. In each panel, the vertical arrows respectively indicate the maximum wavenumber below which a percent-level agreement with N-body simulation is achieved with Lagrangian resummation theory [25,48] and closure theory [22,29], including the PT corrections up to two-loop order.

poor convergence of standard PT expansion, since the low-k behavior of regularized propagators heavily relies on the standard PT treatment. To be specific, the convergence of $\Gamma^{(1)}_{\mathrm{reg}}$ is the main source of this discrepancy. Indeed, if $\Gamma^{(1)}_{\mathrm{reg}}$ is computed at one-loop order only, the power spectrum is enhanced, and then N-body results at low k lie in between the two predictions. The impact of the high-order PT corrections to the two-point propagator are specifically studied in a separate publication, [38].

(ii) Another discrepancy can be found in the high-z results, which temporally overshoot the N-body results at mid-k regime ($k \sim 0.2$–$0.3h$ Mpc^{-1}). It is unlikely to be due to a poor convergence of standard PT expansion. We rather think that the performances of the N-body simulations might be responsible for this (small) discrepancy. We have tested several runs with different resolutions, and found that the low-resolution simulation with a small number of particles tends to underestimate the power at high z. Possible reason for this comes from the precision of force calculation around the intervening scales, where the tree and particle-mesh algorithms are switched, and we suspect that the discrepancy is mainly attributed to the inaccuracy of the tree algorithm. Though the intervening scale is usually set at a sufficiently small scale, with a low-resolution simulation, it may affect the large-scale dynamics with noticeable effects at higher redshifts. Systematic studies on the convergence and resolution of N-body simulations will be reported elsewhere [42].

Apart from the tiny systematics at subpercent level, REGPT approach can give a reliable power spectrum prediction at rather wider range, which entirely covers the relevant scales of BAOs at $z \gtrsim 0.35$. As we will see later in Sec. VI B, the applicable range of the REGPT calculation remains wide enough even in other cosmological models, and can be empirically described with the criterion (42).

C. Correlation function

We next consider the two-point correlation function, which can be computed from the power spectrum as

$$\xi(r) = \int \frac{dkk^2}{2\pi^2} P(k) \frac{\sin(kr)}{kr}. \tag{29}$$

In Fig. 10, left panel focuses on the behaviors around the baryon acoustic peak, while right panel shows the global shape of the two-point correlation function plotted in logarithmic scales, for which $\xi(r)$ has been multiplied by the

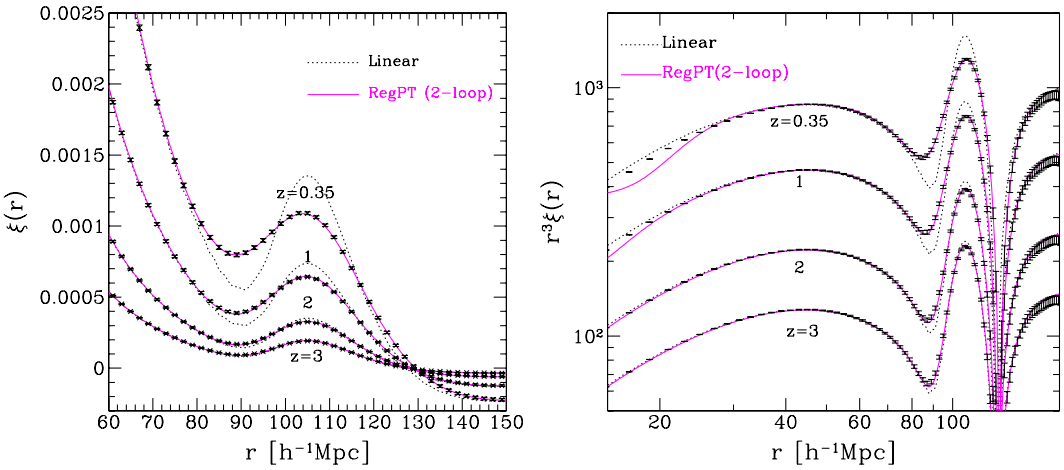


FIG. 10 (color online). Comparison of two-point correlation function between N-body and REGPT results at $z = 3$, 2, 1, and 0.35 (from bottom to top). In each panel, magenta solid, and black dotted lines represent the prediction from REGPT and linear theory calculations, respectively. Left panel focuses on the behavior around baryon acoustic peak in linear scales, while right panel shows the overall behavior in a wide range of separation in logarithmic scales. Note that in right panel, the resulting correlation function is multiplied by the cube of the separation for illustrative purpose.

cube of the separation. The REGPT results agree with N-body simulations almost perfectly over the plotted scales. As it is known, the impact of nonlinear clustering on the baryon acoustic peak is significant: the peak position becomes slightly shifted to a smaller scale, and the structure of the peak tends to be smeared as the redshift decreases (e.g., Refs. [24,25,49,50]). The REGPT calculation can describe not only the behavior around the baryon acoustic peak but also the small-scale behavior of the correlation function. Note that similar results are also obtained from other improved PT treatments such as closure and LRT. Although the REGPT predictions eventually deviate from simulations at small scales—the result at $z = 0.35$ indeed manifests the discrepancy below $r \sim 30h^{-1}$ Mpc—the actual range of agreement between REGPT and N-body results is even wider than what is naively expected from the power spectrum results. In fact, it has been recently advocated by several authors that with several improved PT treatments, the one-loop calculation is sufficient to accurately describe the two-point correlation function (e.g., Refs. [22,48,51]). We have checked that the REGPT treatment at one-loop order can give a satisfactory result close to the two-loop result, and the prediction including the two-loop corrections only slightly improves the agreement with N-body simulations at small scales. This is good news for practical purposes in the sense that we do not necessarily have to evaluate the multidimensional integrals for the accurate prediction of two-point correlation function in the weakly nonlinear regime. Nevertheless, in this work, we keep the two-loop contributions in the computed contributions. The computational costs of the two-loop order will be addressed in the following with the development of a method for accelerated PT calculation at two-loop order.

V. REGPT-FAST: ACCELERATED POWER SPECTRUM CALCULATION

In this section, we present a method that allows accelerated calculations of the required diagrams of the two-loop order REGPT prescription. In principle, the power spectra calculations in the context of REGPT require multidimensional integrations that cannot be done beforehand as they fully depend on the linear power spectra. It is however possible to obtain the required quantities much more rapidly provided we know the answer for a close enough model.

The key point in this approach is to utilize the fact that the nonlinear REGPT power spectrum is a well-defined functional form of the linear power spectrum. Each of the diagrams that has to be computed is of quadratic, cubic, etc. order with respect to the linear power spectrum with a kernel that, although complicated, can be explicitly given. It is then easy to Taylor-expand each of these terms with respect to the linear power spectrum. In principle one then just needs to prepare, in advance, a set of the REGPT results for some fiducial cosmological models, and then take the difference between fiducial and target initial power spectra for which we want to calculate the nonlinear power spectrum. These differences involve only one-dimensional integrals at the first order in the Taylor expansion.

In the following, we present the detail of the implementation of this approach illustrating it with the one-loop calculation case.

A. Power spectrum reconstruction from fiducial model

While our final goal is to present the fast PT calculation at two-loop order, in order to get insights into the implementation of this calculation, we consider the power

spectrum at one-loop order. The complete expressions needed for the fast PT calculation at two-loop order, together with the prescription how to implement it, is presented in Appendix A.

Compared to the expressions given in Eq. (23), the power spectrum at one-loop order of the Γ expansion reduces to

$$P_{ab}(k;\eta) = \Gamma^{(1)}_{a,\mathrm{reg}}(k;\eta)\Gamma^{(1)}_{b,\mathrm{reg}}(k;\eta)P_0(k) + 2\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}\Gamma^{(2)}_{a,\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta) \times \Gamma^{(2)}_{b,\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)P_0(q)P_0(|\boldsymbol{k}-\boldsymbol{q}|) \quad (30)$$

with the regularized propagators $\Gamma^{(1)}_{\mathrm{reg}}$ and $\Gamma^{(2)}_{\mathrm{reg}}$ valid at one-loop order being

$$\Gamma^{(1)}_{a,\mathrm{reg}}(k;\eta) = e^{\eta}\left[1+\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}+e^{2\eta}\bar{\Gamma}^{(1)}_{a,1-\mathrm{loop}}(k)\right] \times \exp\left\{-\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\right\}, \quad (31)$$

$$\Gamma^{(2)}_{a,\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta) = e^{2\eta}F^{(2)}_{a,\mathrm{sym}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})\exp\left\{-\frac{k^2\sigma_{\mathrm{d}}^2e^{2\eta}}{2}\right\}. \quad (32)$$

Note that the quantity $\bar{\Gamma}^{(1)}_{a,1\text{-loop}}$ is defined in Eq. (19), and explicitly given by

$$\bar{\Gamma}^{(1)}_{a,1-\mathrm{loop}}(k) = 3\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}F^{(3)}_{a,\mathrm{sym}}(\boldsymbol{q},-\boldsymbol{q},\boldsymbol{k})P_0(q). \quad (33)$$

Thus, in Eq. (30), there apparently appear two contributions which involve multidimensional integrals; $\bar{\Gamma}^{(1)}_{a,1\text{-loop}}$ in the regularized propagator $\Gamma^{(1)}_{a,\mathrm{reg}}$, and the second term at the right-hand side. Although these contributions are known to be further reduced to one- and two-dimensional integrals (e.g., Refs. [23,29,37]), respectively, for the sake of this presentation we keep the expressions as in their original form.

As has been mentioned earlier, the key idea of accelerated calculation is to prepare a set of REGPT results for fiducial cosmological models. Let us denote the initial power spectrum for fiducial cosmology by $P_{0,\mathrm{fid}}(k)$. And we denote the initial spectrum for the target cosmological model, for which we want to calculate the nonlinear power spectrum, by $P_{0,\mathrm{target}}(k)$. For the moment, we assume that the difference between those spectra is small enough. Then, we may write

$$P_{0,\mathrm{target}}(k) = P_{0,\mathrm{fid}}(k) + \delta P_0(k). \quad (34)$$

Hereafter, we focus on the power spectrum of density field, P_{11}, and drop the subscript. Substituting the above expression into Eqs. (30)–(32), the nonlinear power spectrum for the target model is symbolically written as

$$P_{\mathrm{target}}(k;\eta) = P_{\mathrm{un-pert}}[k,\eta,\sigma_{\mathrm{d,target}};P_{0,\mathrm{fid}}(k)] + P_{\mathrm{corr}}[k,\eta,\sigma_{\mathrm{d,target}};\delta P_0(k)]. \quad (35)$$

Here, the first term at the right-hand side is the unperturbed part of the one-loop power spectrum, which is nothing but the expression (30) adopting the initial power spectrum for fiducial model, $P_{0,\mathrm{fid}}(k)$, but with the cosmological dependence of the time variable, given by $\eta = \ln D(z)$, being calculated from the target model. Also, the dispersion of displacement field, σ_{d}, should be replaced with the one for the target model, i.e., $\sigma_{\mathrm{d,target}} = [\int dq P_{0,\mathrm{target}}(q)/(6\pi^2)]^{1/2}$. In each term of Eq. (30), the scale and time dependence can be separately treated, and thus the unperturbed power spectrum, $P_{\mathrm{un\text{-}perturbed}}$, is evaluated algebraically by summing up each contribution, for which we use the precomputed data set in evaluating the scale-dependent function.

In Eq. (35), the contribution P_{corr} includes the nonlinear corrections originating from the differences of initial power spectra between fiducial and target cosmological models. To first order in δP_0, we have

$$P_{\mathrm{corr}}[k,\eta,\sigma_{\mathrm{d,target}};\delta P_0(k)] = 2\Gamma^{(1)}_{\mathrm{reg}}(k;\eta)\delta\Gamma^{(1)}_{\mathrm{reg}}(k;\eta)P_{0,\mathrm{fid}}(k) + [\Gamma^{(1)}_{\mathrm{reg}}(k;\eta)]^2\delta P_0(k) + 4\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}[\Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)]^2P_{0,\mathrm{fid}}(|\boldsymbol{k}-\boldsymbol{q}|)\delta P_0(q). \quad (36)$$

In the above expression, The quantity σ_{d} appearing in the propagators $\Gamma^{(1)}_{\mathrm{reg}}$ and $\Gamma^{(2)}_{\mathrm{reg}}$ should be evaluated with the linear power spectrum for the target cosmological model. The perturbed propagator $\delta\Gamma^{(1)}_{\mathrm{reg}}$ is expressed as

$$\delta\Gamma^{(1)}_{\mathrm{reg}}(k;\eta) = e^{3\eta}\delta\bar{\Gamma}^{(1)}(k)e^{-k^2\sigma^2_{\mathrm{d,target}}e^{2\eta}/2}, \quad (37)$$

where the kernel of integral in $\delta\bar{\Gamma}^{(1)}$ is the same one as in Eq. (33), but we may rewrite it with

$$\delta\bar{\Gamma}^{(1)}(k) = \int\frac{dqq^2}{2\pi^2}L^{(1)}_1(q,k)\delta P_0(q) \quad (38)$$

with the kernel $L^{(1)}_1$ given by

$$L^{(1)}_1(q,k) = 3\int\frac{d^2_q\boldsymbol{\Omega}_q}{4\pi}F^{(3)}_{1,\mathrm{sym}}(\boldsymbol{q},-\boldsymbol{q},\boldsymbol{k}). \quad (39)$$

Since the kernel $L^{(1)}_1$ only includes the PT kernel whose cosmological dependence is extremely weak, we can

TABLE II. Cosmological parameters for fiducial models used for the REGPT-FAST calculation (see Sec. VI A).

Name	$\Omega_{\rm m}$	Ω_Λ	$\Omega_{\rm b}/\Omega_{\rm m}$	w	h	n_s	σ_8
WMAP3	0.234	0.766	0.175	-1	0.734	0.961	0.760
M001	0.430_7	0.569_2	0.150	-0.816	0.597_7	0.946_8	0.816_1
M023	0.160_2	0.839_8	0.181_7	-1.261	0.869_4	0.901_6	0.666_4

separately prepare the numerical data set for $L_1^{(1)}$ in advance.[10] Then, we can use it to compute $\delta\bar{\Gamma}^{(1)}$ for arbitrary δP_0, where the remaining integral to be evaluated is reduced to a one-dimensional integral.

Furthermore, the integral in the last term of Eq. (36) is rewritten with

$$\int \frac{d^3\boldsymbol{q}}{(2\pi)^3}[\Gamma^{(2)}_{\rm reg}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)]^2 P_{0,\rm fid}(|\boldsymbol{k}-\boldsymbol{q}|)\delta P_0(q) = e^{-k^2\sigma^2_{\rm d,target}e^{2\eta}}e^{4\eta}\int\frac{dqq^2}{2\pi^2}X^{(2)}(q,k)\delta P_0(q) \tag{40}$$

with the function $X^{(2)}$ being

$$X^{(2)}(q,k) = \frac{1}{2}\int_{-1}^{1} d\mu[F^{(2)}_{\rm sym}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})]^2 \times P_{0,\rm fid}\left(\sqrt{k^2-2kq\mu+q^2}\right), \tag{41}$$

where the variable μ is the directional cosine defined by $\mu = (\boldsymbol{k}\cdot\boldsymbol{q})/(kq)$. In deriving the above expression, we used the symmetric property of $\Gamma^{(2)}_{\rm reg}$, i.e., $\Gamma^{(2)}_{\rm reg}(\boldsymbol{k}_1,\boldsymbol{k}_2) = \Gamma^{(2)}_{\rm reg}(\boldsymbol{k}_2,\boldsymbol{k}_1)$. Since the quantity $X^{(2)}(q,k)$ can be computed in advance, all the integrals involving the power spectrum δP_0 are shown to be effectively reduced to one-dimensional integrals. In other words the only remaining task is to evaluate one-dimensional integrals, which can be done very efficiently.

The practical implementation of this method makes use of another important property of the kernel functions. They indeed have a very simple dependence on a global rescaling of the power spectrum, $P_{0,\rm fid} \to cP_{0,\rm fid}$. It is then possible, without extra numerical computation, to choose the fiducial model among a continuous set of models. The model we choose, that is the normalization factor c we take, is such that the difference $\delta P_0(k)$ is as small as possible in the wave-modes of interest. As we will see in Sec. VI it makes the use of this method very efficient.

Note that although the treatment depicted here does not give much impact on the computational cost of the one-loop calculation, we will explicitly show in Appendix A that at two-loop order the PT corrections involving multidimensional integrals can be similarly reduced to one-dimensional integrals. In the following, we denote REGPT-FAST the implementation of this approach at two-loop order.

[10]Indeed, the kernel $L_1^{(1)}$ is analytically known, and the explicit expression is given in, e.g., Refs. [23,29,37].

B. Performances

Let us now illustrate the efficiency of the REGPT-FAST expansion. Based on the expressions given in Appendix A, we calculate the power spectrum and correlation function at two-loop order. We adopt the best-fit parameters determined by the third-year WMAP result [52] as the fiducial cosmological model from which we try to reproduce the REGPT results for the five-year WMAP cosmological model. Cosmological parameters for the fiducial model is listed in Table II. Compared to the target model in Table I, the mass density parameter shows a 20% difference, and with 7% enhancement in the power spectrum normalization (σ_8), this leads to a 20–30% difference in the initial power spectrum.

Figure 11 plots the results of the REGPT-FAST calculation (blue) compared to the target REGPT calculation (magenta). We plot, for a specific redshift $z = 1$, the ratio of the power spectrum to the smooth reference spectrum, $P(k)/P_{\text{no-wiggle}}(k)$, and correlation function multiplied by the cube of separation, $r^3\xi(r)$, in left and right panels, respectively. The REGPT-FAST results perfectly coincide with REGPT direct calculation, even outside the range of agreement with N-body simulations.

Note that the perfect match between REGPT and REGPT-FAST results is due to a large extent to the contributions of the higher-order PT in the correction, $P_{\rm corr}$ or $\xi_{\rm corr}$. This appears clearly in the plots of the linear theory correction, $\delta P_0 = P_{0,\rm target} - P_{0,\rm fid}$ and its Fourier counterpart $\delta\xi_0$ (cyan long-dashed). As shown in cyan solid lines, the total contribution, i.e., the combination of the unperturbed part plus linear theory correction, somehow resembles the result with direct REGPT calculation, but exhibits a rather prominent oscillatory feature with slightly different phase in power spectrum, leading to a non-negligible discrepancy. Accordingly, in correlation function, the acoustic peak becomes enhanced, and the position of peak is shifted to a small separation. Note finally, that these results could only be achieved with the help of the rescaling properties of the kernel functions. In this particular case the fiducial model has been rescaled as $P_{0,\rm fid} \to 1.082P_{0,\rm fid}$. Rescaling is a key feature of the REGPT-FAST method. It will be further discussed in the next section.

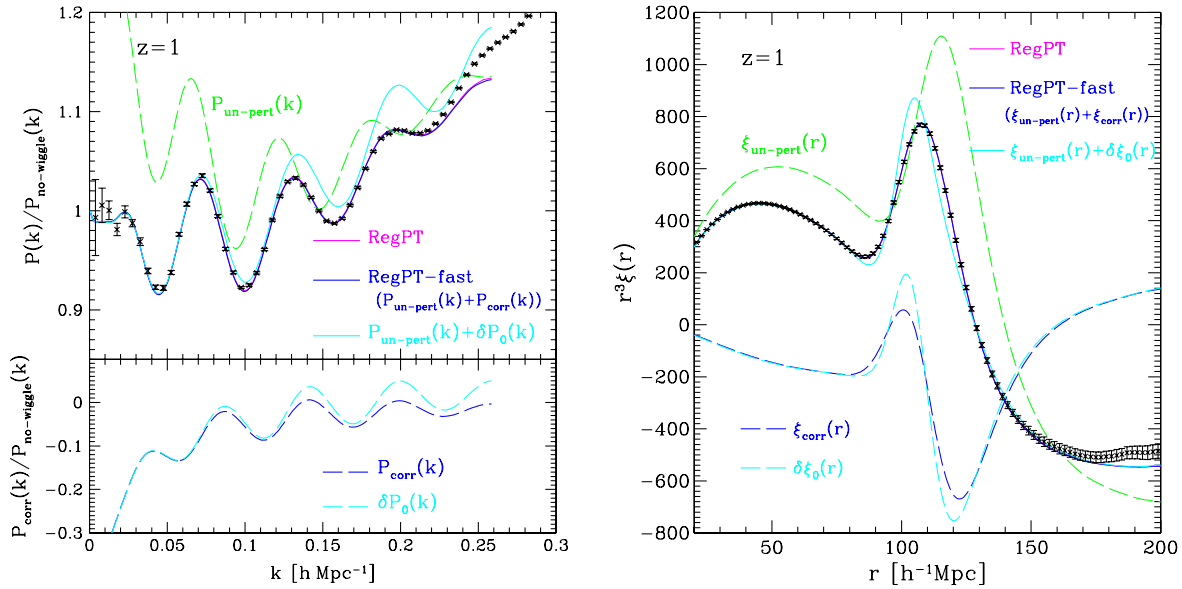


FIG. 11 (color online). Example of the performances of the REGPT-FAST approach compared to direct REGPT calculation. Left panel shows the power spectrum divided by the smooth reference spectrum, $P(k)/P_{\rm no-wiggle}(k)$, while right panel plots the correlation function multiplied by the cube of separation, i.e., $r^3\xi(r)$. In both panels, the results at $z = 1$ are shown, together with N-body simulations. The REGPT-FAST results, computed with prepared data set for fiducial cosmological model, are plotted as magenta solid lines, which almost coincide with those obtained from the rigorous REGPT calculation (solid magenta). As shown in Eq. (35), the REGPT-FAST results are divided into two contributions; unperturbed part ($P_{\rm un-pert}$ or $\xi_{\rm un-pert}$) adopting the WMAP3 model as fiducial cosmology, and the correction part ($P_{\rm corr}$ or $\xi_{\rm corr}$) evaluated with the power spectrum difference δP_0. These are respectively plotted as green dashed and cyan dashed lines. For reference, we also show the linearly evolved result of power spectrum difference δP_0 (dashed cyan) and the sum of the contributions $P_{\rm un-pert} + \delta P_0$ (solid cyan) in left panel, and their Fourier counterparts in right panel.

VI. TESTING REGPT TREATMENT FOR VARYING COSMOLOGICAL MODELS

The purpose of this section is twofold. Our first goal is to explore the validity and applicability of the REGPT-FAST scheme. Having shown that the REGPT-FAST approach can be used in one specific example, we now want to discuss the usefulness of this treatment from a more practical point of view. To be precise, we want to know how well the REGPT-FAST treatment can reproduce rigorous REGPT calculation in a variety of cosmological models.

Our second and natural goal is to test the REGPT scheme itself, whether from direct or fast calculations, against N-body based predictions such that the COSMIC EMULATOR.[11]

To do that, we have selected the 38 cosmological models investigated in Ref. [53] for which we can use the publicly released code, COSMIC EMULATOR, that provides interpolated power spectra derived from N-body simulations. Let us remind that the cosmological models considered there are sampled from a wide parameter space for flat wCDM cosmology, and lie within the range

[11]http://www.lanl.gov/projects/cosmology/CosmicEmu/.

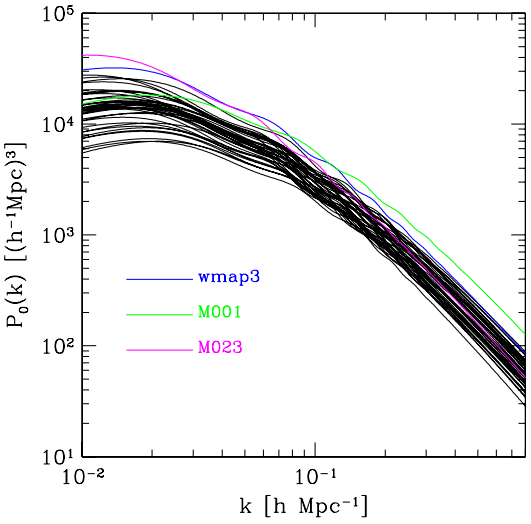


FIG. 12 (color online). Linear power spectra $P_0(k)$ for 38 cosmological models [53]. Blue, green, and magenta lines are respectively the power spectra of fiducial models WMAP3, M001, and M023 used for the REGPT-FAST calculation (see Table II for their cosmological parameters).

$$0.120 < \Omega_m h^2 < 0.155, \qquad 0.0215 < \Omega_b h^2 < 0.0235,$$
$$0.85 < n_s < 1.05, \qquad -1.30 < w < -0.70,$$
$$0.616 < \sigma_8 < 0.9.$$

The concrete values of the cosmological parameters in each model are not shown here. Readers can find them in Table 1 of Ref. [53]. Figure 12 shows the linear power spectra $P_0(k)$ for the 38 cosmological models, which have been all produced with CMB Boltzmann code, CAMB [54].

A. Convergence of REGPT-FAST

Let us first examine the convergence of the power spectrum calculations between REGPT and REGPT-FAST treatments. We ran both the REGPT-FAST and REGPT codes, and evaluated the fractional difference between these power spectra, defined by $P_{\rm RegPTfast}(k)/P_{\rm RegPT}(k) - 1$. Collecting the results at $z = 1$ in each cosmological model, the convergence of the power spectrum calculations for 38 models is summarized in Fig. 13. In left panel we show the result when only one fiducial model, WMAP3, is used. In that case REGPT-FAST results tend to underestimate the results from rigorous REGPT calculations at increasing k, and most of them eventually exceed the 1% difference, indicated by the green dashed line. This is because the shape of the initial power spectrum in each target model is rather different from that in the fiducial model, and even adjusting the rescaling parameter c cannot compensate a large power spectrum difference. To be more precise, in most of the models, the shape parameter, defined by $\Gamma = \Omega_m h$, is typically larger than the one in the fiducial model. As a consequence, even if we adjust the power spectrum at large scales to match the one in the target model, the difference $|\delta P_0|$ can become large as increasing k, leading to a failure of the perturbative reconstruction by REGPT-FAST.

To remedy this situation, a simple but efficient approach is to enlarge our set of fiducial models with Γ parameters that differ from the one of WMAP3 model, i.e., $\Gamma = 0.172$. Right panel of Fig. 13 shows the convergence results when we supply two extra fiducial models whose cosmological parameters are listed in Table II. As a fiducial model with a larger shape parameter, we adopt the M001 cosmological model ($\Gamma = 0.257$). Further, for a secure calculation applicable to general cosmological models, we also supply another fiducial model, M023, which has a smaller shape parameter ($\Gamma = 0.139$). The initial power spectra of those models are plotted in Fig. 12, depicted as green (M001) and magenta (M023) solid lines. As a result, the convergence of the power spectrum calculations is dramatically improved, and the REGPT-FAST now coincides with rigorous REGPT calculation with $\lesssim 0.4\%$ precision at $k \lesssim 0.3h$ Mpc^{-1}. Although there still exist exceptional cases, in which the fractional difference eventually exceeds 1% precision at $k \gtrsim 0.36h$ Mpc^{-1}, in practice this is beyond the applicable range of the REGPT calculation itself.

With this setting, making use of these three fiducial models, REGPT-FAST reproduces REGPT direct calculations in a wide range of cosmological models and, also it does not appear here, for a redshift range of general interest.

B. Comparison with cosmic emulator

It is now time to discuss the accuracy of the overall REGPT scheme with general COSMIC EMULATOR predictions. Figures 14 and 15 summarize the results of the comparison for all 38 models, where we plot the ratios of power spectra, $P(k)/P_{\text{no-wiggle}}(k)$, at specific redshift $z = 1$. In each panel, magenta solid and black dashed lines represent the results of REGPT-FAST and the power spectrum emulator code, respectively. Also, the fiducial model used for the REGPT-FAST calculation is indicated, together with the label of the cosmological model. The two results mostly coincide with each other, and are hardly distinguishable at $k \lesssim 0.2h$ Mpc^{-1}, where the linear theory prediction typically produces a 10% error. At $k \gtrsim 0.2h$ Mpc^{-1}, the REGPT-FAST results tend to deviate from the predictions of the emulator code which probably indicates the the limitation of PT treatment. However, some models still show a remarkable agreement at $k \lesssim 0.3h$ Mpc^{-1} (e.g., M009 and M013).

As the range of applicability of the REGPT scheme depend on both k and the power spectrum amplitude, following Refs. [22,55], we propose here a phenomenological rule for the domain of applicability of the REGPT calculations. The proposed upper value for k is $k_{\rm crit}$ that can be obtained from the implicit equation

$$\frac{k_{\rm crit}^2}{6\pi^2}\int_0^{k_{\rm crit}} dq P_{\rm lin}(q;z) = C, \tag{42}$$

where C is a fixed constant, $C = 0.7$. The resulting values are depicted as vertical arrows in Figs. 14 and 15. Below the critical wavenumber, the REGPT scheme indeed agrees with results of the emulator code, mostly within a percent-level precision.[12] We have also checked that this is also the case for $z = 0.5$ with this definition of $k_{\rm crit}$.

[12]We however noticed that some models exhibit non-negligible discrepancy between the results of REGPT-FAST and the emulator codes, even well below $k_{\rm crit}$. One such is M015, showing a broad-band discrepancy over the plotted range. This is somewhat surprising in the sense that the REGPT-FAST result almost converges the linear theory prediction at $k \lesssim 0.12h$ Mpc^{-1}, while the result of the emulator code is still away from it. To better understand the source of the discrepancy, we have ran N-body simulations for the M015 model—cosmological parameters of M015 model were set as $\Omega_m = 0.2364$, $\Omega_b = 0.0384$, $w = -1.281$, $h = 0.7737$, $n_s = 1.0177$, and $\sigma_8 = 0.7692$.—with the same setup as listed in Table I. The resulting power spectrum, estimated from the ensemble of the 8 independent realizations, is superposed in the panel of M015 (green symbols with errorbars) and is shown to faithfully trace the REGPT-FAST result up to the critical wavenumber. It points to a possible flaw in the power spectrum emulator, in estimating the smooth power spectrum from the ensemble of simulation results, or constructing the interpolated result of the simulated power spectra.

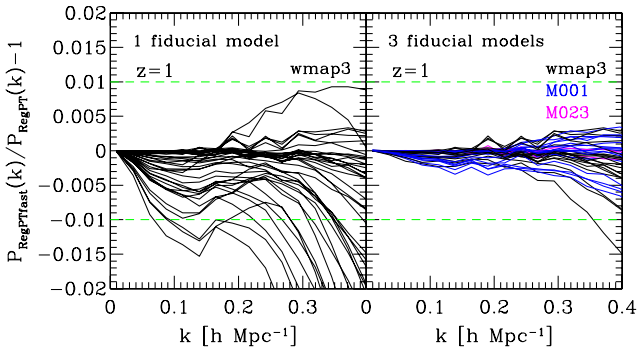


FIG. 13 (color online). Convergence between the REGPT-FAST treatment and the direct REGPT calculations for 38 cosmological models. The fractional difference of the power spectra between REGPT-FAST and REGPT direct calculations, $P_{\mathrm{RegPTfast}}(k)/P_{\mathrm{RegPT}}(k)-1$, is plotted at $z=1$. Left and right panels respectively show the results adopting the one and three fiducial models.

The REGPT scheme is therefore shown to give a fairly accurate prediction for the power spectrum in the weakly nonlinear regime in the sense given above. REGPT direct calculations, or (almost) equivalently, REGPT-FAST calculations with the three fiducial models we prepared, can be applied to a wide range of cosmological models. Though we did not discuss it here, we expect the same to be also true for the correlation function. Finally we note that as the relevant scale of weakly nonlinear regime grows wider for higher redshifts, the applicability and reliability of the REGPT scheme is naturally enhanced. On the other hand, the emulation schemes to build up interpolated results from large sets of N-body simulations are generally efficient in predicting the power spectrum at nonlinear scales but are more likely to fail at high-z, since the requirement for the force resolution in N-body simulation becomes more and more severe. In this respect, perturbative reconstruction

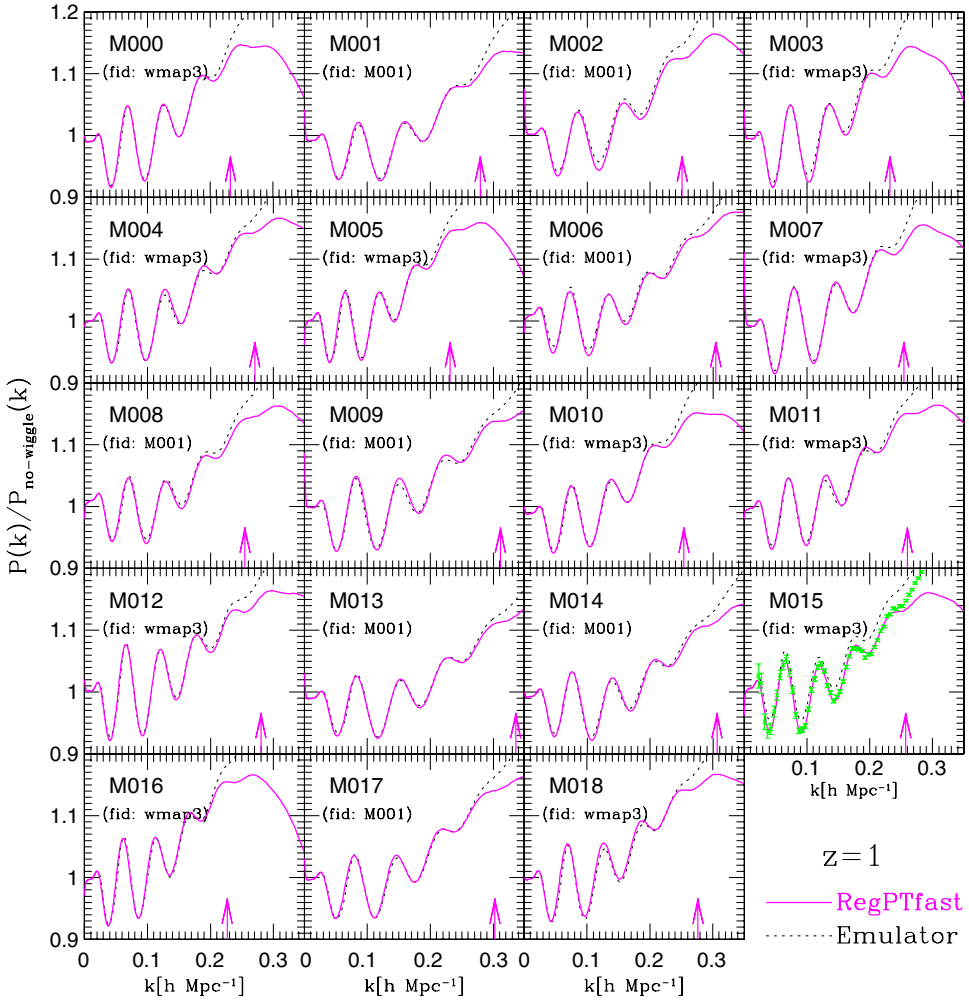


FIG. 14 (color online). Ratio of power spectra, $P(k)/P_{\mathrm{no-wiggle}}(k)$, at $z=1$ for the cosmological models M000–M017. Solid and dotted lines are obtained from the REGPT-FAST and COSMIC EMULATOR codes, respectively. The fiducial model used for the REGPT-FAST calculation is indicated in each panel. The vertical arrows mean the critical wavenumber k_{crit} defined by Eq. (42), which roughly gives an applicable range of REGPT prediction (see text).

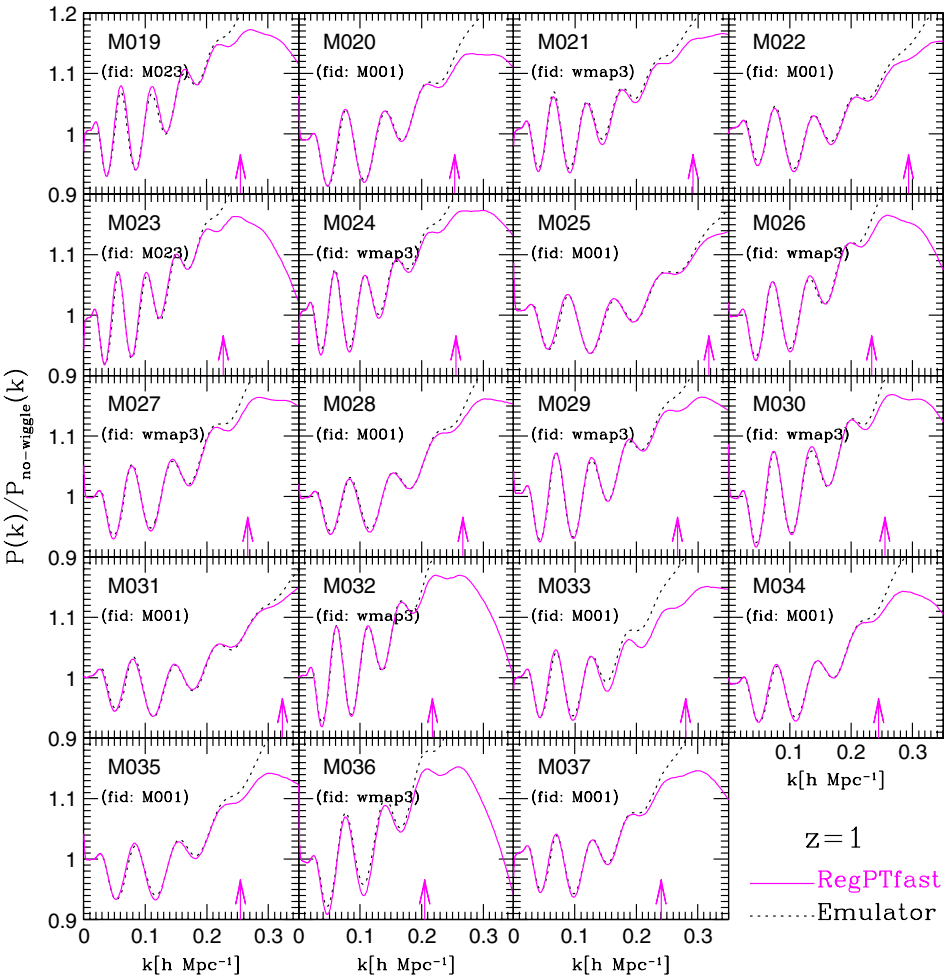


FIG. 15 (color online). Same as Fig. 14, but for the models M019–M037.

schemes such as REGPT-FAST—but this would also be the case of MPTBREEZE—are complementary to N-body based predictions.

VII. CONCLUSION

It is needless to say that future cosmological observations make the development of cosmological tool aiming at accurately predicting the large-scale statistical properties of the universe highly desirable. In the first part of the present paper, based on a renormalized perturbation theory, we introduced an explicit computation scheme applied to the matter power spectrum and correlation function in weakly nonlinear regime that consistently includes the PT corrections up to the two-loop order. The construction of the full expression for the power spectrum is based on the Γ expansion, i.e., makes use of the multipoint propagators which are properly regularized so as to recover their expected resummed behavior at high k and to match the standard PT result at low k. We call this regularized PT treatment REGPT. We have shown that the REGPT scheme provides an accurate prediction for both the power spectrum and the correlation function, leading to a percent-level agreement with N-body simulations in the weakly nonlinear regime.

In the second half of the paper, we presented a method to accelerate the power spectrum calculations. The method utilizes prepared data sets for some specific fiducial models from which regularized PT calculations can be performed for arbitrary cosmological models. The main interest of this method is that the evaluation of the residual PT corrections between fiducial and target cosmological models can be reduced to mere one-dimensional integrals. This enables us to dramatically reduce the computational cost, and even with single-node calculation by a laptop computer, the power spectrum calculation can be done in a few seconds. We call this method REGPT-FAST, and we have demonstrated that the REGPT-FAST treatment can perfectly reproduce the direct REGPT calculations that involve several multidimensional integrals.

We then investigated the range of applicability of the REGPT schemes in a broad class of cosmological models. For this purpose, we select 38 cosmological models, and compared the REGPT predictions—eventually incorporating

the accelerated computations—with results of a power spectrum emulator code, COSMIC EMULATOR. We show that with the help of three fiducial models the REGPT-FAST calculations give reliable predictions for the power spectra over this range of cosmological models.[13] We furthermore put forward an empirical criterion (42) that gives a good indication of the applicable range of the REGPT scheme in k. This can be applied to any cosmological model, and we found that the applicable range of REGPT scheme remains fairly wide in a wide range of cosmological models, entirely covering the relevant scales of BAOs. The REGPT-FAST treatment, together with the direct REGPT calculation, has been implemented in a Fortran code that we publicly release as part of this paper.

Although this paper has focused on precision calculations of the matter power spectrum, the REGPT framework as well as the methodology for accelerated calculation can naturally be applied to the power spectrum of the velocity divergence and the cross-power spectrum of velocity and density fields in a similar way. The analysis of the velocity power spectrum, together with a detailed comparison with N-body simulations, will be presented elsewhere. Of particular interest is the application of the REGPT schemes to the redshift-space power spectrum or correlation function. In this case, not only the velocity and density power spectra, but also the multipoint spectra like bispectrum, arising from the nonlinear mode coupling, seem to play important roles, and should be properly modeled. Significance of the effect of multipoint spectra has been recently advocated by Refs. [51,56–58], and there appear physical models that account for this. Combination of these models with the REGPT schemes would be very important, and we will discuss it in a near future.

ACKNOWLEDGMENTS

A. T. is grateful to Toshiya Namikawa for his helpful comments on the code implentation. S. C. thanks Christophe Pichon for fruitful comments and discussion. This work has been benefited from exchange visits supported by a bilateral grant from Ministère Affaires Etrangères et Européennes in France and Japan Society for the Promotion of Science (JSPS). A. T. acknowledges support from Institutional Program for Young Researcher Overseas Visit funded by the JSPS. A. T. is also supported in part by a Grant-in-Aid for Scientific Research from the JSPS (Grant No. 24540257). T. N. is supported by a Grant-in-Aid for JSPS Fellows (Grant No. PD: 22-181) and by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan. Numerical computations for the present work have been carried out in part on Cray XT4 at Center for Computational Astrophysics, CfCA, of National Astronomical Observatory of Japan, and in part under the Interdisciplinary Computational Science Program in Center for Computational Sciences, University of Tsukuba. F. B and S. C. are also partly supported by the French Programme National de Cosmologie et Galaxies.

APPENDIX A: PERTURBATIVE RECONSTRUCTION OF REGPT POWER SPECTRUM AT TWO-LOOP ORDER

In this appendix, we present the set of perturbative expressions that are used for the accelerated power spectrum calculation at two-loop order which is implemented in the REGPT-FAST code.

In a similar manner to the one-loop case described in Sec. V A, we can expand the power spectrum expression up to two-loop order around the fiducial cosmological model, and obtain the perturbative expression for power spectrum in the target cosmological model. Plugging Eq. (34) into the two-loop expression (23) and assuming $\delta P_0 \ll P_{0,\mathrm{fid}}$, the power spectrum is written like (35), and the correction P_{corr} becomes

$$\begin{aligned} P_{\mathrm{corr}}[k,\eta,\sigma_{\mathrm{d,target}};\delta P_0(k)] &= 2\Gamma^{(1)}_{\mathrm{reg}}(k;\eta)\delta\Gamma^{(1)}_{\mathrm{reg}}(k;\eta)P_0(k) + [\Gamma^{(1)}_{\mathrm{reg}}(k;\eta)]^2\delta P_0(k) + 4\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}\{[\Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)]^2 \\ &\quad\times P_0(|\boldsymbol{k}-\boldsymbol{q}|)\delta P_0(q) + \Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)\delta\Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)P_0(|\boldsymbol{k}-\boldsymbol{q}|)P_0(q)\} \\ &\quad + 18\int\frac{d^3\boldsymbol{p}d^3\boldsymbol{q}}{(2\pi)^6}[\Gamma^{(3)}_{\mathrm{reg}}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q};\eta)]^2P_0(|\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}|)P_0(p)\delta P_0(q). \end{aligned} \tag{A1}$$

In the above, the perturbations of regularized propagators, $\delta\Gamma^{(1)}_{\mathrm{reg}}$ and $\delta\Gamma^{(2)}_{\mathrm{reg}}$, are described as

$$\begin{aligned} \delta\Gamma^{(1)}_{\mathrm{reg}}(k;\eta) &= e^{3\eta}[(1+\alpha_k)\delta\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k) \\ &\quad + e^{2\eta}\delta\bar{\Gamma}^{(1)}_{2-\mathrm{loop}}(k)]e^{-\alpha_k}, \end{aligned} \tag{A2}$$

$$\delta\Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta) = e^{4\eta}\delta\bar{\Gamma}^{(2)}_{1-\mathrm{loop}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})e^{-\alpha_k}, \tag{A3}$$

where we define $\alpha_k \equiv k^2\sigma^2_{\mathrm{d,target}}e^{2\eta}/2$. The quantities $\delta\bar{\Gamma}^{(p)}_{\mathrm{n-loop}}$ are defined by

$$\delta\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k) = 3\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}F^{(3)}_{\mathrm{sym}}(\boldsymbol{q},-\boldsymbol{q},\boldsymbol{k})\delta P_0(q), \tag{A4}$$

[13]Our analysis is however restricted to flat wCDM models.

$$\delta\bar{\Gamma}^{(1)}_{2-\mathrm{loop}}(k) = 30\int\frac{d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2}{(2\pi)^6}F^{(5)}_{\mathrm{sym}}(\boldsymbol{q}_1,-\boldsymbol{q}_1,\boldsymbol{q}_2,-\boldsymbol{q}_2,\boldsymbol{k}) \times P_{0,\mathrm{fid}}(q_1)\delta P_0(q_2), \tag{A5}$$

$$\delta\bar{\Gamma}^{(2)}_{1-\mathrm{loop}}(\boldsymbol{k}_1,\boldsymbol{k}_2) = 6\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}F^{(4)}_{\mathrm{sym}}(\boldsymbol{q},-\boldsymbol{q},\boldsymbol{k}_1,\boldsymbol{k}_2)\delta P_0(q). \tag{A6}$$

The kernels $F^{(p)}_{\mathrm{sym}}$ are the symmetrized standard PT kernel for density field. In the above, the angular integrals are known to be analytically performed (Refs. [23,37], and Bernardeau *et al.* in preparation), one may write

$$\delta\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k) = \int\frac{dqq^2}{2\pi^2}f(q;k)\delta P_0(q), \tag{A7}$$

$$\delta\bar{\Gamma}^{(1)}_{2-\mathrm{loop}}(k) = 2\int\frac{dq_1dq_2q_1^2q_2^2}{(2\pi^2)^2}J(q_1,q_2;k)P_{0,\mathrm{fid}}(q_1)\delta P_0(q_2), \tag{A8}$$

$$\delta\bar{\Gamma}^{(2)}_{1-\mathrm{loop}}(\boldsymbol{k}_1,\boldsymbol{k}_2) = \int\frac{dqq^2}{2\pi^2}K(q;k_1,k_2,k_3)\delta P_0(q) \tag{A9}$$

with the angle-averaged kernels f, J and K defined by

$$f(q;k) = 3\int\frac{d^2\boldsymbol{\Omega}_q}{4\pi}F^{(3)}_{\mathrm{sym}}(\boldsymbol{q},-\boldsymbol{q},\boldsymbol{k}), \tag{A10}$$

$$J(q_1,q_2;k) = 15\int\frac{d^2\boldsymbol{\Omega}_{q1}d^2\boldsymbol{\Omega}_{q_2}}{(4\pi)^2}F^{(5)}_{\mathrm{sym}}(\boldsymbol{q}_1,-\boldsymbol{q}_1,\boldsymbol{q}_2,-\boldsymbol{q}_2,\boldsymbol{k}), \tag{A11}$$

$$K(q;k_1,k_2,k_3) = 6\int\frac{d^2\boldsymbol{\Omega}_q}{4\pi}F^{(4)}_{\mathrm{sym}}(\boldsymbol{q},-\boldsymbol{q},\boldsymbol{k}_1,\boldsymbol{k}_2). \tag{A12}$$

Note that $\boldsymbol{k}_1+\boldsymbol{k}_2=\boldsymbol{k}_3$.

The expression for the correction P_{corr} given above contains many integrals involving the perturbed linear power spectrum, δP_0, and some of these require multi-dimensional integrals. However, those multidimensional integration are separately treated, and can be effectively reduced to the one-dimensional integrals as follows,

$$\delta\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k) = \int\frac{dqq^2}{2\pi^2}L^{(1)}(q,k)\delta P_0(q), \tag{A13}$$

$$\delta\bar{\Gamma}^{(1)}_{2-\mathrm{loop}}(k) = 2\int\frac{dqq^2}{2\pi^2}M^{(1)}(q,k)\delta P_0(q), \tag{A14}$$

$$\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}[\Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)]^2P_{0,\mathrm{fid}}(|\boldsymbol{k}-\boldsymbol{q}|)\delta P_0(q) = e^{4\eta}\Big[(1+\alpha_k)^2\int\frac{dqq^2}{2\pi^2}X^{(2)}(q,k)\delta P_0(q) + 2e^{2\eta}(1+\alpha_k)\int\frac{dqq^2}{2\pi^2}Y^{(2)}(q,k)\delta P_0(q) + e^{4\eta}\int\frac{dqq^2}{2\pi^2}Z^{(2)}(q,k)\delta P_0(q)\Big]\exp\{-2\alpha_k\}, \tag{A15}$$

$$\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}\Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta)\delta\Gamma^{(2)}_{\mathrm{reg}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q};\eta) \times P_{0,\mathrm{fid}}(|\boldsymbol{k}-\boldsymbol{q}|)P_{0,\mathrm{fid}}(q) = e^{6\eta}\Big[(1+\alpha_k)\int\frac{dpp^2}{2\pi^2}Q^{(2)}(p,k)\delta P_0(p) + e^{2\eta}\int\frac{dpp^2}{2\pi^2}R^{(2)}(p,k)\delta P_0(p)\Big]e^{-2\alpha_k}, \tag{A16}$$

$$\int\frac{d^3\boldsymbol{p}d^3\boldsymbol{q}}{(2\pi)^6}[\Gamma^{(3)}_{\mathrm{reg}}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q};\eta)]^2 \times P_{0,\mathrm{fid}}(|\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}|)P_{0,\mathrm{fid}}(p)\delta P_0(q) = e^{6\eta}e^{-2\alpha_k}\int\frac{dqq^2}{2\pi^2}S^{(3)}(q,k)\delta P_0(q). \tag{A17}$$

In the above, the kernels of the integrals, L, M, X, Y, Z, Q, R, and S, additionally need to be computed, but we only have to evaluate them once for each fiducial cosmological model,

$$L^{(1)}(q,k) = f(q;k), \tag{A18}$$

$$M^{(1)}(q,k) = \int\frac{dpp^2}{2\pi^2}J(p,q;k)P_{0,\mathrm{fid}}(p), \tag{A19}$$

$$X^{(2)}(q,k) = \frac{1}{2}\int_{-1}^{1}d\mu_q[F^{(2)}_{\mathrm{sym}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})]^2 \times P_{0,\mathrm{fid}}\Big(\sqrt{k^2-2kq\mu_q+q^2}\Big), \tag{A20}$$

$$Y^{(2)}(q,k) = \frac{1}{2}\int_{-1}^{1}d\mu_qF^{(2)}_{\mathrm{sym}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})\bar{\Gamma}^{(2)}_{1-\mathrm{loop}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q}) \times P_{0,\mathrm{fid}}\Big(\sqrt{k^2-2kq\mu_q+q^2}\Big), \tag{A21}$$

$$Z^{(2)}(q,k) = \frac{1}{2}\int_{-1}^{1}d\mu_q[\bar{\Gamma}^{(2)}_{1-\mathrm{loop}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})]^2 \times P_{0,\mathrm{fid}}\Big(\sqrt{k^2-2kq\mu_q+q^2}\Big), \tag{A22}$$

$$Q^{(2)}(p,k) = \int\frac{d^3\boldsymbol{q}}{(2\pi)^3}F^{(2)}_{\mathrm{sym}}(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q})K(p\colon q,|\boldsymbol{k}-\boldsymbol{q}|,k) \times P_{0,\mathrm{fid}}(|\boldsymbol{k}-\boldsymbol{q}|)P_{0,\mathrm{fid}}(q), \tag{A23}$$

$$R^{(2)}(p,k) = \int \frac{d^3\boldsymbol{q}}{(2\pi)^3} \bar{\Gamma}^{(2)}_{1-\mathrm{loop}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) K(p; q, |\boldsymbol{k}-\boldsymbol{q}|, k) \times P_{0,\mathrm{fid}}(|\boldsymbol{k}-\boldsymbol{q}|) P_{0,\mathrm{fid}}(q), \tag{A24}$$

$$S^{(3)}(q,k) = \frac{1}{2}\int_{-1}^{1} d\mu_q \int \frac{d^3\boldsymbol{p}}{(2\pi)^3} [F^{(3)}_{\mathrm{sym}}(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q})]^2 \times P_{0,\mathrm{fid}}(|\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}|) P_{0,\mathrm{fid}}(p), \tag{A25}$$

with the variable μ_q defined by $\mu_q = (\boldsymbol{k}\cdot\boldsymbol{q})/(kq)$.

Note that similar to the one-loop case, the correction at two-loop order also possesses a one-parameter degree of freedom corresponding to a global rescaling of the power spectrum of fiducial model, $P_{0,\mathrm{fid}} \to cP_{0,\mathrm{fid}}$. The power spectrum difference δP_0 can then be made small securing the efficient convergence of this expansion.

Finally, the data set of kernel functions given above are supplemented in the REGPT code, with 301×301 logarithmic arrays in (k, q) space. For specific three fiducial models (i.e., WMAP3, M001, and M023), the data have been obtained using the method of Gaussian quadrature up to three-dimensional integrals and Monte Carlo technique for four-dimensional integral. Together with unperturbed part of the PT corrections, these can be used as fast calculations of power spectrum at two-loop order.

APPENDIX B: CODE DESCRIPTION

In this appendix, we present a detailed description of the Fortran code, REGPT, which computes the power spectrum and correlation function of density fields valid at weakly nonlinear regime of gravitational clustering.

1. Overview

The code, REGPT, is compiled with the Fortran compilers, IFORT or GFORTRAN. It computes the power spectrum in flat wCDM class models based on the REGPT treatment when provided with either of transfer function or matter power spectrum. It then gives the multiple-redshift outputs for power spectrum, and optionally provides correlation function data. We have implemented two major options for power spectrum calculations:

- `-fast`: Applying the reconstruction method described in Sec. VA, this option quickly computes the power spectrum at two-loop level (typically a few seconds), using the pre-computed data set of PT kernels for fiducial cosmological models. We provide the data set for three fiducial models (WMAP3, M001, and M023, see Table II), and the code automatically finds an appropriate fiducial model to closely match the result of rigorous PT calculation with direct-mode.
- `-direct`: With this option, the code first applies the fast method, and then follows the regularized expression for power spectrum (see Eq. (23) with regularized propagators (24)–(26)) to directly evaluate the multidimensional integrals (it typically takes a few minutes). The output results are the power spectrum of direct calculation and difference of the results between fast and direct method. Further, the code gives the data set of PT diagrams necessary for power spectrum calculations, from which we can construct the power spectrum. We provide a supplemental code, `read_stfile.f`, with which the power spectrum and correlation function can be evaluated from the diagram data set in several PT methods, including the standard PT and LRT [25,48] as well as RegPT treatment (see Appendix B 4 c).

In addition, the code supports the option, `-direct1loop`, to compute the power spectrum at one-loop order. Although this is based on the direct calculation with multidimensional integration (see Eq. (30) with regularized propagators (31) and (32)), the one-loop expression involves two-dimensional integrals at most, and thus the computational cost is less expensive. It is potentially useful for the computation of high-z correlation function and power spectrum.

2. Setup

The REGPT code is available at Ref. [59]. A part of REGPT code uses the library for Monte Carlo integration, CUBA [40]. Before compiling the codes, users should download the library package `cuba-1.5`, and correctly build the file, `libcuba.a`, compatible with the architecture of user's platform. This can be done in the directory `/Cuba-1.5`, and just type "`./configure`" and "`makelib`." After placing the library file `libcuba.a` at the directory `/RegPT/src`, users can use the `Makefile` to create the main executable file, `RegPT.exe`. Note that currently available compilers are Intel Fortran compiler, `ifort`, and GNU Fortran compiler, `gfortran`.

3. Running the code

Provided with linear power spectrum or transfer function data, the code runs with a set of options, and computes power spectrum. Users can specify the options in the command line, or using the parameter file (suffix of file name should be `.ini`). Sample of parameter file is supplied in the code (see directory `/RegPT/example`).

For running the code with the command-line options, a simple example is (assuming the code is placed at the directory, `/RegPT`)

```
./RegPT.exe␣-spectrum␣-infile␣matterpower_wmap5.dat␣-nz␣2␣0.5␣1.0
```

In the above example, the code first reads the input data file, `matterpower_wmap5.dat`, which is assumed to contain linear power spectrum data consisting of two columns, i.e., k and $P_0(k)$. By default setting, `fast` mode is chosen, and the output result of power spectrum is saved to `pk_RegPT.dat`. With the option `-nz 2 0.5 1.0`, the output file contains the power spectrum results at two redshifts, $z = 0.5$ and 1.0 (see Appendix B 4 a for output format). Note that by default, the code adopts specific values of cosmological parameters. Making use of options, users can change the value of cosmological parameters appropriately, consistently with input power spectrum (or transfer function) data.

Here we summarize the available options to run the code:

- Verbose level for output message
  `-verbose` n: This sets the verbose level for output information on the progress of numerical computation. The available level n is 1 or 2 (default: `-verbose1`).
  `-noverbose`: This option suppresses the message while running the code.
- Input data file
  `-infile` [file]: Input file name of power spectrum or transfer function data is specified (default: `-infile matterpower.dat`).
  `-path` [path to input file]: This specifies the path to the input file (default: `-path./`).
  `-spectrum`: With this option, the code assumes that the input file is power spectrum data. The data consists of two columns, i.e., wavenumber (in units of h Mpc^{-1}) and matter power spectrum (in units of h^{-3} Mpc^3) (default: `-spectrum`). The normalization of power spectrum amplitude can be made with the option `-sigma8`.
  `-transfer`: With this option, the code assumes that the input file is the transfer function data created by CAMB. The data should contain seven columns, among which the code uses the first and seven columns (wavenumber in units of h Mpc^{-1} and matter transfer function). The normalization of power spectrum amplitude can be made with either of the option `-sigma8` or `-samp` and `-spivot`.
- Specification of cosmological parameters
  `-sigma8` σ_8: This option sets the power spectrum normalization by σ_8 (default: `-sigma8 0.817`). For $\sigma_8 < 0$, the code will skip the σ_8 normalization.
  `-samp` A_s: This option sets the amplitude of power spectrum at pivot scale $k_{\rm pivot}$ (default: `-samp 2.1e-9`). This option is used for normalization of transfer function data, and is valid when the option `-transfer` is specified.
  `-spivot` $k_{\rm pivot}$: This option sets the pivot scale of CMB normalization in units of Mpc^{-1} (default: `-spivot 0.05`). This option is used for normalization of transfer function data, and is valid when the option `-transfer` is specified.
  `-omegam` $\Omega_{\rm m}$: This option sets the mass density parameter (default: `-omegam 0.279`). This is used to estimate the linear growth factor and to compute the smooth reference spectrum, $P_{\rm no\text{-}wiggle}(k)$.
  `-omegab` $\Omega_{\rm b}$: This option sets the baryon density parameter (default: `-omegab 0.165 * omegam`). This is used to compute the smooth reference spectrum, $P_{\rm no\text{-}wiggle}(k)$.
  `-ns` n_s: This option sets the scalar spectral index. This is used to compute the linear power spectrum from the transfer function data (option `-transfer` should be specified), and to compute the smooth reference spectrum, $P_{\rm no\text{-}wiggle}(k)$.
  `-w` w: This option sets the equation of state for dark energy (default: `-w -1.0`). This is used to estimate the linear growth factor.
  `-h` h: This option sets the Hubble parameter (default: `-h 0.701`). This is used to compute the power spectrum from the transfer function data, and to compute the smooth reference spectrum, $P_{\rm no\text{-}wiggle}(k)$.
  `-camb` [output parameter file of camb]: With this option, the code reads the CAMB output parameter file, and specifies the cosmological parameters ($\Omega_{\rm m}$, $\Omega_{\rm b}$, w, h, n_s, A_s, $k_{\rm pivot}$).
- Calculation mode of REGPT
  `-fast`: This option adopts the fast method of power spectrum calculation to give REGPT results. This is default setting.
  `-direct`: This option first applies the fast method, and then follow the direct method for REGPT calculation.
  `-direct1loop`: With this option, the code adopts direct method to compute the power spectrum at one-loop order.
- Setup of fiducial models for fast- and direct-mode calculations
  `-datapath` [path to data directory]: This option specifies the path to the data files used for power spectrum calculation with fast and direct methods (default: `-datapath data/`). In the directory specified with this option, the data set of kernel functions given in Appendix A and unperturbed part of power spectrum corrections, as well as the matter power spectrum should be stored for three fiducial cosmological models (WMAP3, M001, and M023).
  `-fiducial` [model]: This option sets the specific fiducial model among the three, WMAP3, M001, and M023 (in default setting, the code automatically selects an appropriate fiducial model).
- Output data file
  `-xicompute`: With this option, the code computes the correlation function after power spectrum calculations, and creates the output file.
  `-nz` $n\, z_1 \cdots z_n$: This option specifies the output redshifts for power spectrum calculations. The integer n

specifies the number of redshifts, and subsequent arguments specify the value of each redshift (default: `-nz 1 1.0`).
`-pkfile` [file]: This option sets the output file name of power spectrum data (default: `pk_RegPT.dat`).
`-xifile` [file]: This option sets the output file name of correlation function data (default: `xi_RegPT.dat`).
`-stfile` [file]: This option sets the output file name of PT diagram data (default: `st_PT.dat`).

4. Output file format

In what follows, wavenumber k and separation r are in units of h Mpc^{-1} and h^{-1} Mpc, respectively. All the power spectrum data are assumed to be in units of h^{-3} Mpc^3.

a. Power spectrum data

By default, REGPT code creates the output file for the power spectrum data (default file name is `pk_RegPT.dat`). The columns of this file include

$$k, [\text{data for } z_1], [\text{data for } z_2], \ldots, [\text{data for } z_n].$$

The first column is the wavenumber, while the bracket [data for z_i] represents a set of power spectra at given redshift z_i and wavenumber k. Number of the data set is specified with the option `-nz`, and each data contains

$$P_{\text{no-wiggle}}(k, z_i), \quad P_{\text{lin}}(k, z_i), \quad P_{\text{RegPT}}(k, z_i), \quad \text{Err}(k).$$

Here, the spectrum $P_{\text{no-wiggle}}$ is the smooth reference spectrum calculated from the no-wiggle formula of linear transfer function in Ref. [47], P_{lin} is the linearly extrapolated spectrum, and $P_{\text{RegPT}}(k, z_i)$ represents the power spectrum based on the REGPT calculations with fast and/or direct method (depending on the choice of options, `-fast`, `-direct` or `-direct1loop`). The last column, Err, usually sets to zero, but with the option `-direct`, it gives the difference of the power spectra between fast and direct methods.

b. Correlation function data

With the option `-xicompute`, the code also provides the output file for correlation function data (default file name is `xi_RegPT.dat`). Similar to the power spectrum data, the structure of the data is

$$r, [\text{data for } z_1], [\text{data for } z_2], \ldots, [\text{data for } z_n]$$

The first column is the separation, while the bracket [data for z_i] represents a set of correlation functions given at redshift z_i and separation r, containing two columns

$$\xi_{\text{lin}}(r, z_i), \qquad \xi_{\text{RegPT}}(r, z_i).$$

These are simply obtained from the output results of power spectrum based on the expression (29). Note that the range of wavenumber for output power spectrum is restricted to the wavenumber coverage of input linear spectrum (or transfer function). To get a convergent result of correlation functions, users may have to supply the input data file with a sufficiently wide range of wavenumber (e.g., $10^{-3} \le k \le 10h$ Mpc^{-1}).

c. Diagram data

When users specify the `-direct` option, the code additionally provides a set of PT diagram data necessary for power spectrum computation, from which we can construct the power spectrum at one- and two-loop order. The output file (default file name is `st_PT.dat`) includes the following columns:

$$k,\ P_{\text{no-wiggle}}(k),\ P_{\text{lin}}(k),\ \bar{\Gamma}^{(1)}_{1\text{-loop}}(k),\ \bar{\Gamma}^{(1)}_{2\text{-loop}}(k),\ P^{(2)\text{tree-tree}}_{\text{corr}}(k),\ P^{(2)\text{tree-1loop}}_{\text{corr}}(k),\ P^{(2)1\text{loop-1loop}}_{\text{corr}}(k),\ P^{(3)\text{tree-tree}}_{\text{corr}}(k).$$

Here, the power spectra $P_{\text{no-wiggle}}$ and P_{lin} are basically the same data as contained in the power spectrum file, but these are the extrapolated data at $z = 0$ (that is, P_{lin} corresponds to P_0). The function $\bar{\Gamma}^{(1)}_{n\text{-loop}}$ is the two-point propagator of the standard PT expansion (see definition [(19)]). The functions in the remaining four columns, $P^{(2)\text{tree-tree}}_{\text{corr}}$, $P^{(2)\text{tree-1loop}}_{\text{corr}}$, $P^{(2)1\text{loop-1loop}}_{\text{corr}}$, and $P^{(3)\text{tree-tree}}_{\text{corr}}$, are defined by

$$P^{(2)\text{tree}-\text{tree}}_{\text{corr}}(k) = 2\int \frac{d^3\boldsymbol{q}}{(2\pi)^3} F^{(2)}_{\text{sym}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) F^{(2)}_{\text{sym}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) \times P_0(q) P_0(|\boldsymbol{k}-\boldsymbol{q}|), \tag{B1}$$

$$P^{(2)\text{tree}-1\text{loop}}_{\text{corr}}(k) = 4\int \frac{d^3\boldsymbol{q}}{(2\pi)^3} F^{(2)}_{\text{sym}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) \bar{\Gamma}^{(2)}_{1-\text{loop}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) \times P_0(q) P_0(|\boldsymbol{k}-\boldsymbol{q}|), \tag{B2}$$

$$P^{(2)1\text{loop}-1\text{loop}}_{\text{corr}}(k) = 2\int \frac{d^3\boldsymbol{q}}{(2\pi)^3} \bar{\Gamma}^{(2)}_{1-\text{loop}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) \times \bar{\Gamma}^{(2)}_{1-\text{loop}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) P_0(q) P_0(|\boldsymbol{k}-\boldsymbol{q}|), \tag{B3}$$

$$P^{(3)\text{tree}-\text{tree}}_{\text{corr}}(k) = 6\int \frac{d^3\boldsymbol{p} d^3\boldsymbol{q}}{(2\pi)^6} F^{(3)}_{\text{sym}}(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}) \times F^{(3)}_{\text{sym}}(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}) P_0(p) P_0(q) \times P_0(|\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}|). \tag{B4}$$

Provided the data set above, the power spectrum can be constructed with

$$P^{\mathrm{RegPT}}_{1-\mathrm{loop}}(k;\eta) = e^{2\eta}e^{-2\alpha_k}[\{1+\alpha_k+e^{2\eta}\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k)\}^2P_0(k)+e^{2\eta}P^{(2)\mathrm{tree}-\mathrm{tree}}_{\mathrm{corr}}(k)], \tag{B5}$$

$$P^{\mathrm{RegPT}}_{2-\mathrm{loop}}(k;\eta) = e^{2\eta}e^{-2\alpha_k}\Bigg[\Bigg\{1+\alpha_k+\frac{\alpha_k^2}{2}+e^{2\eta}\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k)(1+\alpha_k)+e^{4\eta}\bar{\Gamma}^{(1)}_{2-\mathrm{loop}}(k)\Bigg\}^2P_0(k)+e^{2\eta}\{(1+\alpha_k)^2P^{(2)\mathrm{tree}-\mathrm{tree}}_{\mathrm{corr}}(k) + e^{2\eta}(1+\alpha_k)P^{(2)\mathrm{tree}-1\mathrm{loop}}_{\mathrm{corr}}(k)+e^{4\eta}P^{(2)1\mathrm{loop}-1\mathrm{loop}}_{\mathrm{corr}}(k)\}+e^{4\eta}P^{(3)1\mathrm{loop}-1\mathrm{loop}}_{\mathrm{corr}}(k)\Bigg] \tag{B6}$$

for the REGPT calculation at one- and two-loop order, respectively. Here, α_k is given by $\alpha_k = k^2\sigma_{\mathrm{d}}^2e^{2\eta}/2$ with σ_{d} being the dispersion of displacement field (see Eq. (16)). Note that the diagram data set can be also used to compute the power spectrum in the standard PT calculations

$$P^{\mathrm{SPT}}_{1-\mathrm{loop}}(k;\eta) = e^{2\eta}P_0(k)+e^{4\eta}[2P_0(k)\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k) + P^{(2)\mathrm{tree}-\mathrm{tree}}_{\mathrm{corr}}(k)], \tag{B7}$$

$$P^{\mathrm{SPT}}_{2-\mathrm{loop}}(k;\eta) = P^{\mathrm{SPT}}_{1-\mathrm{loop}}(k;\eta)+e^{6\eta}[P_0(k)\{\bar{\Gamma}^{(1)}_{1-\mathrm{loop}}(k)\}^2 + P^{(3)\mathrm{tree}-\mathrm{tree}}_{\mathrm{corr}}(k)+P^{(2)\mathrm{tree}-1\mathrm{loop}}_{\mathrm{corr}}(k) + 2P_0(k)\bar{\Gamma}^{(1)}_{2-\mathrm{loop}}(k)]. \tag{B8}$$

With the supplemental code, `read_stfile.f`, users can easily compute the power spectrum in both REGPT and standard PT treatments. The code also provides the power spectrum result for LRT [25,48]. A brief instruction on how to run the code and the output format of data is described in the header of the code.

5. Limitation

Since the REGPT code is the PT-based calculation code valid at weakly nonlinear scales, the applicability of the output results is restricted to a certain range of wavenumber in power spectrum. We provide an empirical estimate of critical wavenumber k_{crit}, below which the REGPT results are reliable and their accuracy can reach a percent level. This is based on Eq. (42) with constant value $C = 0.7(0.3)$ for two-loop (one-loop) (see Sec. VI B). With the option `-verbose 2`, the code displays the critical wavenumbers at output redshifts. Note that the value k_{crit} given here is just a crude estimate, and the actual domain of applicability may be somewhat wider or narrower. Users should use the output results with a great care.

[1] D. J. Eisenstein *et al.* (SDSS Collaboration), Astrophys. J. **633**, 560 (2005).
[2] W. J. Percival *et al.*, Mon. Not. R. Astron. Soc. **401**, 2148 (2010).
[3] C. Blake, T. Davis, G. Poole, D. Parkinson, S. Brough *et al.*, Mon. Not. R. Astron. Soc. **415**, 2892 (2011).
[4] H.-J. Seo, S. Ho, M. White, A. Cuesta, A. Ross *et al.*, arXiv:1201.2172.
[5] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton *et al.*, arXiv:1203.6594.
[6] H.-J. Seo and D. J. Eisenstein, Astrophys. J. **598**, 720 (2003).
[7] C. Blake and K. Glazebrook, Astrophys. J. **594**, 665 (2003).
[8] K. Glazebrook and C. Blake, Astrophys. J. **631**, 1 (2005).
[9] M. Shoji, D. Jeong, and E. Komatsu, Astrophys. J. **693**, 1404 (2009).
[10] N. Padmanabhan and M. J. White, Phys. Rev. D **77**, 123 540 (2008).
[11] E. V. Linder, Astropart. Phys. **29**, 336 (2008).
[12] L. Guzzo *et al.*, Nature **451**, 541 (2008).
[13] K. Yamamoto, T. Sato, and G. Huetsi, Prog. Theor. Phys. **120**, 609 (2008).
[14] Y.-S. Song and W. J. Percival, J. Cosmol. Astropart. Phys. 10 (2009) 004.
[15] C. Blake, S. Brough, M. Colless, C. Contreras, W. Couch *et al.*, Mon. Not. R. Astron. Soc. **415**, 2876 (2011).
[16] L. Van Waerbeke, Y. Mellier, T. Erben, J. C. Cuillandre, F. Bernardeau, R. Maoli, E. Bertin, H. J. McCracken, O. Le Fèvre, B. Fort *et al.*, Astron. Astrophys. **358**, 30 (2000).
[17] L. Fu, E. Semboloni, H. Hoekstra, M. Kilbinger, L. van Waerbeke, I. Tereno, Y. Mellier, C. Heymans, J. Coupon, K. Benabed *et al.*, Astron. Astrophys. **479**, 9 (2008).
[18] M. Bartelmann and P. Schneider, Phys. Rep. **340**, 291 (2001).
[19] L. Van Waerbeke and Y. Mellier, arXiv:astro-ph/0305089.
[20] M. Crocce and R. Scoccimarro, Phys. Rev. D **73**, 063 519 (2006).
[21] J. Carlson, M. White, and N. Padmanabhan, Phys. Rev. D **80**, 043 531 (2009).
[22] A. Taruya, T. Nishimichi, S. Saito, and T. Hiramatsu, Phys. Rev. D **80**, 123 503 (2009).
[23] M. Crocce and R. Scoccimarro, Phys. Rev. D **73**, 063 520 (2006).
[24] M. Crocce and R. Scoccimarro, Phys. Rev. D **77**, 023 533 (2008).
[25] T. Matsubara, Phys. Rev. D **77**, 063 530 (2008).
[26] T. Matsubara, Phys. Rev. D **78**, 083 519 (2008).
[27] P. McDonald, Phys. Rev. D **75**, 043 514 (2007).

[28] K. Izumi and J. Soda, Phys. Rev. D **76**, 083 517 (2007).
[29] A. Taruya and T. Hiramatsu, Astrophys. J. **674**, 617 (2008).
[30] M. Pietroni, J. Cosmol. Astropart. Phys. 10 (2008) 036.
[31] S. Matarrese and M. Pietroni, J. Cosmol. Astropart. Phys. 06 (2007) 026.
[32] P. Valageas, Astron. Astrophys. **421**, 23 (2004).
[33] P. Valageas, Astron. Astrophys. **465**, 725 (2007).
[34] F. Bernardeau, M. Crocce, and R. Scoccimarro, Phys. Rev. D **78**, 103 521 (2008).
[35] F. Bernardeau, M. Crocce, and E. Sefusatti, Phys. Rev. D **82**, 083 507 (2010).
[36] F. Bernardeau, N. Van de Rijt, and F. Vernizzi, Phys. Rev. D **85**, 063 509 (2012).
[37] F. Bernardeau, M. Crocce, and R. Scoccimarro, Phys. Rev. D **85**, 123 519 (2012).
[38] F. Bernardeau, A. Taruya, and T. Nishimichi (to be published).
[39] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002).
[40] T. Hahn, Comput. Phys. Commun. **168**, 78 (2005).
[41] M. Crocce, R. Scoccimarro, and F. Bernardeau, arXiv:1207.1465.
[42] T. Nishimichi, A. Taruya, and F. Bernardeau (to be published).
[43] V. Springel, Mon. Not. R. Astron. Soc. **364**, 1105 (2005).
[44] E. Komatsu *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **180**, 330 (2009).
[45] M. Crocce, S. Pueblas, and R. Scoccimarro, Mon. Not. R. Astron. Soc. **373**, 369 (2006).
[46] P. Valageas and T. Nishimichi, Astron. Astrophys. **527**, A87 (2011).
[47] D. J. Eisenstein and W. Hu, Astrophys. J. **496**, 605 (1998).
[48] T. Okamura, A. Taruya, and T. Matsubara, J. Cosmol. Astropart. Phys. 08 (2011) 012.
[49] D. J. Eisenstein, H.-j. Seo, and M. J. White, Astrophys. J. **664**, 660 (2007).
[50] R. E. Smith, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **77**, 043 525 (2008).
[51] B. A. Reid and M. White, Mon. Not. R. Astron. Soc. **417**, 1913 (2011).
[52] D. N. Spergel *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **170**, 377 (2007).
[53] E. Lawrence, K. Heitmann, M. White, D. Higdon, C. Wagner, S. Habib, and B. Williams, Astrophys. J. **713**, 1322 (2010).
[54] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000).
[55] T. Nishimichi *et al.*, Publ. Astron. Soc. Jpn. **61**, 321 (2009).
[56] A. Taruya, T. Nishimichi, and S. Saito, Phys. Rev. D **82**, 063 522 (2010).
[57] T. Nishimichi and A. Taruya, Phys. Rev. D **84**, 043 526 (2011).
[58] J. Tang, I. Kayo, and M. Takada, Mon. Not. R. Astron. Soc. **416**, 2291 (2011).
[59] http://www-utap.phys.s.u-tokyo.ac.jp/~ataruya/regpt_code.html.

Compte-Rendu Français

Nous proposons ici une description en langue française des principaux résultats présentés dans ce manuscrit de thèse.

B.1 Introduction

Les objets astrophysiques que nous observons sur la voûte céleste ne sont pas répartis uniformément. En parcourant le ciel, notre regard se pose sur diverses agglomérations d'étoiles et sur une gigantesque fresque lumineuse qui traverse le ciel: notre galaxie, la Voie Lactée. Les astronomes ont révélé au monde l'existence d'une multitude d'autres galaxies qui elles-même ne sont pas agencées de façon uniforme dans l'espace mais se rassemblent en des amas et des super-amas. Jusqu'à la fin du siècle dernier, on pense que ces amas sont les plus grandes structures que le cosmos ait vu naître et qu'ils ne suivent aucun motif particulier. Ainsi, le principe cosmologique sur lequel repose notre modélisation de l'Univers suppose que celui-ci est homogène et isotrope mais cette hypothèse semble valide seulement aux plus grandes échelles. En effet, il apparaît dans les années 80 que les galaxies ne sont pas de simples îles voguant uniformément dans l'espace mais forment un véritable réseau en bulles de savon sur des centaines de mega-parsecs! La dynamique de la matière à ces échelles est désormais bien connue en particulier grâce aux grands relevés de galaxies et aux simulations numériques. La matière s'échappe des vides vers les murs puis glisse vers les filaments avant de rejoindre les noeuds sur-denses de cette toile cosmique (voir la figure B.1).

Les galaxies naissent et grandissent au sein même de ce grand ballet cosmique ce qui soulève la question du rôle des grandes structures dans les processus de formation des galaxies et en particulier dans le modelage de leur forme. Leur masse par exemple dépend grandement de l'environnement à grande échelle comme l'explique la théorie de l'effondrement gravitationnel biaisé (biased clustering en anglais) selon laquelle les objets les plus massifs se forment de préférence dans les régions sur-denses telles que les filaments et plus encore les noeuds. Au-delà de leur masse, la question est ensuite de savoir comment les autres propriétés des galaxies sont affectées par l'environnement, en particulier leur moment angulaire, si important pour comprendre la mise en place et l'évolution de leur morphologie (elliptique, spirale) et donc la séquence de Hubble. Une question qui se pose est donc de comprendre comment les axes de rotation des galaxies sont corrélés à la direction des grands filaments de la toile cosmique. Toutes ces questions sont cruciales si l'on souhaite comprendre la formation des galaxies et en particulier le rôle de l'inné – les processus internes – sur l'acquis – l'environnement – (nature versus nurture).

Les grandes structures représentent également un enjeu majeur de la cosmologie. En effet, leur croissance au cours du temps est directement sensible aux propriétés de l'Univers. Les étudier pourrait donc permettre de répondre aux grandes questions soulevées par la cosmologie moderne: quelles sont les lois qui régissent l'Univers? Einstein avait-il vu juste ou sommes-nous capables

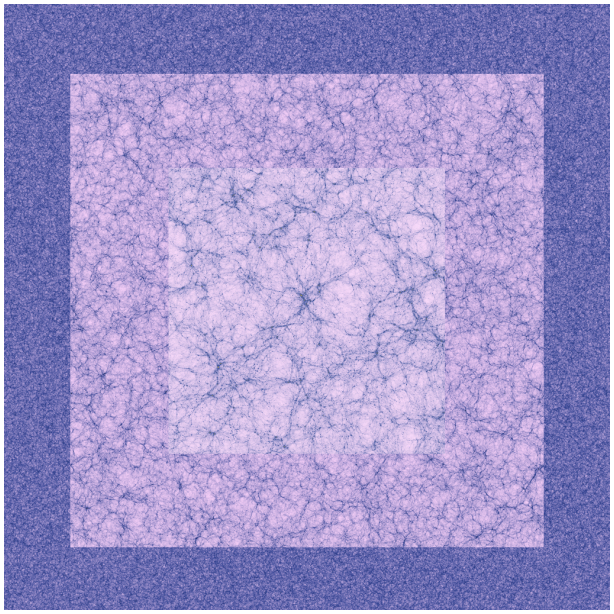


Figure B.1: La simulation Horizon 4π . La région extérieure correspond à une vue de l'Univers sur des échelles de $16h^{-1}$Gpc. La région intermédiaire correspond à une tranche de $2h^{-1}$Gpc tandis que la région interne est à la résolution des conditions initiales. Credits: The Horizon project.

de mesurer des déviations à sa théorie de la Gravitation à savoir le Relativité Générale? Quels sont les constituants du cosmos? Quel est le moteur de l'accélération de l'expansion cosmique?

Pour toutes les raisons mentionnées ci-avant, l'étude des grandes structures de l'Univers est devenue un champs de recherche très actif en astrophysique et est promise à un bel avenir avec notamment l'arrivée de grands relevés comme Euclid ou le LSST qui vont sonder l'Univers à grande échelle et apporter une gigantesque quantité de données à la communauté, l'opportunité d'en savoir un peu plus sur notre monde et peut-être de soulever de nouvelles questions.

Cette thèse a pour but d'exposer les travaux de recherche que j'ai menés au cours de mon doctorat à l'Institut d'Astrophysique de Paris sous la direction de Christophe Pichon et Dmitry Pogosyan entre septembre 2011 et septembre 2015. Ces recherches s'articulent autour de la problématique d'une description théorique des grandes structures et de leur rôle d'une part en cosmologie et d'autre part pour comprendre la formation des galaxies. Ce manuscrit rassemble plus précisément des résultats obtenus en cherchant à répondre à deux questions fondamentales:

- comment utiliser efficacement les grandes structures pour en apprendre davantage sur le cosmos?
- quel est l'impact de la toile cosmique sur la formation des galaxies et en particulier sur leur moment angulaire intrinsèque (spin)?

Nous allons ici donner un compte-rendu en langue française des principaux résultats présentés dans ce manuscrit. Dans une première partie, nous décrirons comment le spin des halos de matière noire ainsi que celui des galaxies est corrélé aux filaments de la toile cosmique obtenue dans des simulations numériques. Ces corrélations peuvent être comprises analytiquement en partant des premiers principes à condition de tenir compte du caractère anisotrope de la toile cosmique dans la théorie standard d'acquisition de spin par effet de marée. Enfin, ces

corrélations entre forme des galaxies et environnement induisent une possible contamination dans les expériences qui utilisent le cisaillement gravitationnel (souvent appelée "alignements intrinsèques"). Cette contamination sera étudiée dans une simulation numérique de pointe – la simulation Horizon-AGN – qui modélise la formation des galaxies dans un volume cosmologique. Dans une seconde partie, nous présenterons des résultats obtenus d'une part sur la topologie des grandes structures en espace des redshifts (décalages vers le rouge) et d'autre part sur les propriétés statistiques du champ de densité dans des sphères concentriques. Ces deux thématiques sont des observables prometteuses pour les années à venir.

B.2 Alignements intrinsèques des galaxies au sein de la toile cosmique

B.2.1 Introduction

L'astigmatisme cosmique ou effet de lentillage gravitationnel faible est souvent présenté comme une prochaine sonde cosmologique majeure avec l'arrivée de grands relevés tels que DES [1], Euclid ([Laureijs et al., 2011]) ou LSST [2]. L'idée est la suivante: parce que la trajectoire des rayons lumineux depuis une source d'arrière-plan jusqu'à nous est courbée par les puits de potentiel gravitationnel sur la ligne de visée, les galaxies que nous observons dans le ciel sont distordues. Ainsi, mesurer ces distorsions permet d'obtenir des informations sur le modèle cosmologique et la distribution de matière noire. Il s'agit donc d'essayer de mesurer des distorsions cohérentes dans les formes des galaxies en utilisant par exemple la fonction de corrélation à deux points. Cependant, il faut ici garder en tête que l'ellipticité apparente d'une galaxie est la somme de deux effets: une distorsion due au cisaillement gravitationnel que nous venons de mentionner mais aussi une composante d'ellipticité intrinsèque $e = e_s + \gamma$, où e est l'ellipticité apparente et e_s l'ellipticité intrinsèque de la source (celle que nous observerions sans effet de lentille). De cette façon, nous pouvons décomposer la fonction de corrélation à deux points de l'ellipticité projetée comme

$$\langle e(\vartheta)e(\vartheta+\theta)\rangle_\vartheta = \langle\gamma\gamma'\rangle + \langle e_s e_s'\rangle + 2\langle e_s\gamma'\rangle \ , \tag{B.1}$$

où à des fins de compacité, le prime signifie à une distance angulaire θ de la première localisation. Les deux derniers termes de l'équation B.1 représentent les deux types d'alignements intrinsèques qui contaminent le signal provenant du cisaillement gravitationnel $\langle\gamma\gamma'\rangle$. Le premier contaminant $\langle e_s e_s'\rangle$ est souvent appelé terme "II" (intrinsèque-intrinsèque) et provient des corrélations intrinsèques des ellipticités dans le plan source ([Heavens et al., 2000, Croft and Metzler, 2000, Catelan et al., 2001]). Le deuxième contaminant $\langle e_s\gamma'\rangle$ est quant à lui appelé terme "GI" (gravitationnel-intrinsèque) et provient des corrélations entre l'ellipticité intrinsèque d'une galaxie et l'ellipticité induite d'une source à plus haut redshift ([Hirata and Seljak, 2004]).

D'un point de vue théorique, de solides arguments permettent de montrer que les formes des galaxies sont corrélées avec la toile cosmique. Cette cohérence à grande échelle devrait donc induire un certain niveau d'alignements intrinsèques et ainsi contaminer les mesures d'astigmatisme cosmique. Nous nous proposons ici de décrire une extension anisotrope de la théorie des champs de marée (tidal torque theory) qui permet de tenir compte de l'effet des grandes structures sur l'acquisition de spin des galaxies. Nous montrerons que les simulations de matière noire pure et les simulations hydrodynamiques (qui modélisent aussi la matière baryonique: gaz, étoiles)

[1] http://www.darkenergysurvey.org
[2] `http://www.lsst.org`

prédisent que les spins sont en effet corrélés à la toile cosmique d'une façon qui dépend de leur masse et du redshift en accord avec la théorie des champs de marée anisotrope susmentionnée. Ces corrélations induisent des alignements intrinsèques qui seront explicitement mesurés dans la simulation hydrodynamique Horizon-AGN .

B.2.2 Spins et grandes structures dans les simulations numériques

Au cours de la dernière décennie, il a été démontré dans des simulations numériques que les halos de matière noire ([Aragón-Calvo et al., 2007b, Paz et al., 2008, Codis et al., 2012] par exemple) et les galaxies ([Hahn et al., 2010, Dubois et al., 2014]) sont corrélés avec la toile cosmique. Dans [Codis et al., 2012], nous montrons que le spin des halos de matière noire est corrélé à la direction du filament le plus proche (voir le panneau de gauche de la figure B.2). Ces résultats sont obtenus à partir des 43 millions de halos de la simulation Horizon 4π et montrent qu'il y a une forte dépendence en masse. Les halos les plus légers tendent à présenter un spin aligné avec l'axe du filament alors que les halos massifs sont plus enclin à avoir un spin perpendiculaire au filament. La transition entre ces deux régimes apparaît pour une masse $M_{\rm tr} \approx 5 \cdot 10^{12} M_\odot$, qui dépend du redshift comme $M_{\rm tr}(z) \approx M_0 (1+z)^{-\gamma_s}$ avec $\gamma_s = 2.5 \pm 0.2$, et qui varie faiblement avec l'échelle du filament.

Ce signal peut être compris dans le contexte de la formation hiérarchique des structures qui voit les objets les moins massifs naître en premier puis former de plus grandes structures par fusions successives. En effet, la première génération de petits halos se forme lorsque les murs s'intersectent pour former des filaments, processus au cours duquel les proto-galaxies vont avoir tendance à acquérir du moment angulaire aligné avec l'axe du filament. Quand elles fusionnent ensuite le long du filament pour former des objets plus massifs, elles convertissent leur moment angulaire orbital en spin perpendiculaire à la direction du mouvement et donc au filament. Une illustration de ce processus est donnée en figure B.3. Dans les deux cas (faible ou forte masse), la clef réside dans le fait que le spin acquis est perpendiculaire à la direction du mouvement qui est le long du mur et perpendiculaire au filament en formation pour les faibles masses et alignée avec le filament pour les fortes masses.

L'émergence d'un champ de vorticité aligné avec le filament et responsable de l'orientation du spin des objets peu massifs a été étudiée dans [Laigle et al., 2015]. Dans cet article, nous avons trouvé que la vorticité est confinée dans les filaments – ce qui est naturel sachant que l'on ne peut générer de la vorticité que dans les régions multi-flots – et qu'elle est statistiquement alignée avec l'axe du filament en accord avec l'idée d'un "enroulement" des murs pour former les filaments. En moyenne, une coupe perpendiculaire au filament est quadrupolaire dans le sens où elle révèle quatre quadrants ayant une orientation de la vorticité alternée (+/-/+/-). Nous montrons aussi que la frontière entre ces quadrants coincide avec la direction des murs (avec un excès de probabilité d'alignement de 15% environ), ce qui renforce l'idée selon laquelle le moteur de la génération de vorticité est bien l'enroulement des murs (voir les panneaux de gauche de la figure B.3). La matière s'échappe des murs en direction des filaments en formation puis coule le long des filaments jusqu'à rejoindre les noeuds. Cette dynamique crée un flot de vorticité parallèle au filament et partitionné en quatre régions ayant une direction différente afin de préserver la symétrie et une vorticité globale nulle. Un filament est ainsi typiquement le croisement de deux murs qui délimitent les frontières de ces quatre régions.

Plus récemment, nous avons étudié dans [Dubois et al., 2014] comment le scénario décrit dans le paragraphe précédent se traduit pour non plus des halos mais des galaxies virtuelles. C'est

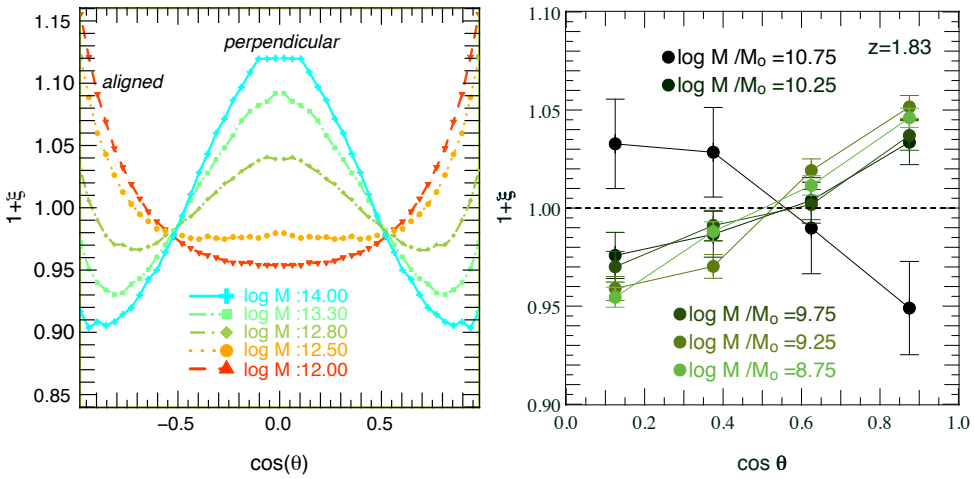


Figure B.2: A gauche: excès de probabilité d'alignement entre spin et direction du filament le plus proche mesuré sur les 43 millions de halos de la simulation Horizon 4π ([Teyssier, 2002]) à redshift zéro. Les différentes couleurs correspondent à des masses de 10^{12}(rouge) à 10^{14} masses solaires (bleu). Une masse de transition est détectée à $M_{\rm crit}(z=0) \simeq 5(\pm 1) \times 10^{12} M_\odot$: les halos de masse $M > M_{\rm crit}$ ont un spin plutôt perpendiculaire à leur filament hôte alors que les halos de plus faible masse $M < M_{\rm crit}$ ont un spin plutôt aligné avec le filament. Cette figure est extraite de [Codis et al., 2012]. A droite: même mesure mais avec les 160 000 galaxies de la simulation Horizon-AGN à redshift $z = 1.8$ ([Dubois et al., 2014]).

dans ce but que nous avons utilisé la simulation Horizon-AGN – une simulation cosmologique basée sur le code hydrodynamique Ramses – et avons trouvé que les spins galactiques étaient eux aussi corrélés avec la direction des filaments. Les galaxies peu massives, bleues, spirales et qui forment des étoiles tendent à avoir un axe de rotation aligné avec l'axe du filament alors que les galaxies massives, rouges, elliptiques et peu actives ont plutôt tendance à avoir un spin perpendiculaire aux filaments. Ce résultat est illustré sur le panneau de droite de la figure B.2. Il faut noter que de récentes observations (en particulier sur le relevé du Sloan Digital Sky Survey, par exemple [Tempel and Libeskind, 2013]) ont rapporté des indices en faveur d'une corrélation entre morphologie des galaxies et toile cosmique comme prédite par les simulations numériques.

B.2.3 Une théorie de l'acquisition de spin au voisinage des filaments cosmiques

Les corrélations entre spin et toile cosmique telles que prédites par les simulations numériques peuvent être comprises dans le cadre de la théorie des champs de marée – théorie selon laquelle les proto-galaxies acquièrent leur spin principalement pendant le régime linéaire de croissance des structures par l'effet du champ de marée de la matière environnante – et ce, à condition de tenir compte de l'influence des filaments à l'échelle galactique. Il s'agit de réaliser que les murs et les filaments agissent comme un mode de grande longueur d'onde par dessus lequel les petites fluctuations, qui vont mener à la formation des halos, naissent. Ce mode à grande échelle induit un boost anisotrope dans le champ de densité permettant aux pics de passer le seuil critique d'effondrement nécessaire pour se découpler de l'expansion globale de l'Univers et ainsi former un objet gravitationnellement lié. C'est cet effet de biais qui est ainsi évoqué pour justifier de la distribution des galaxies au voisinage proche des noeuds sur-denses de la toile cosmique ([White et al., 1988]). Cela explique pourquoi les galaxies se forment préférentiellement dans les

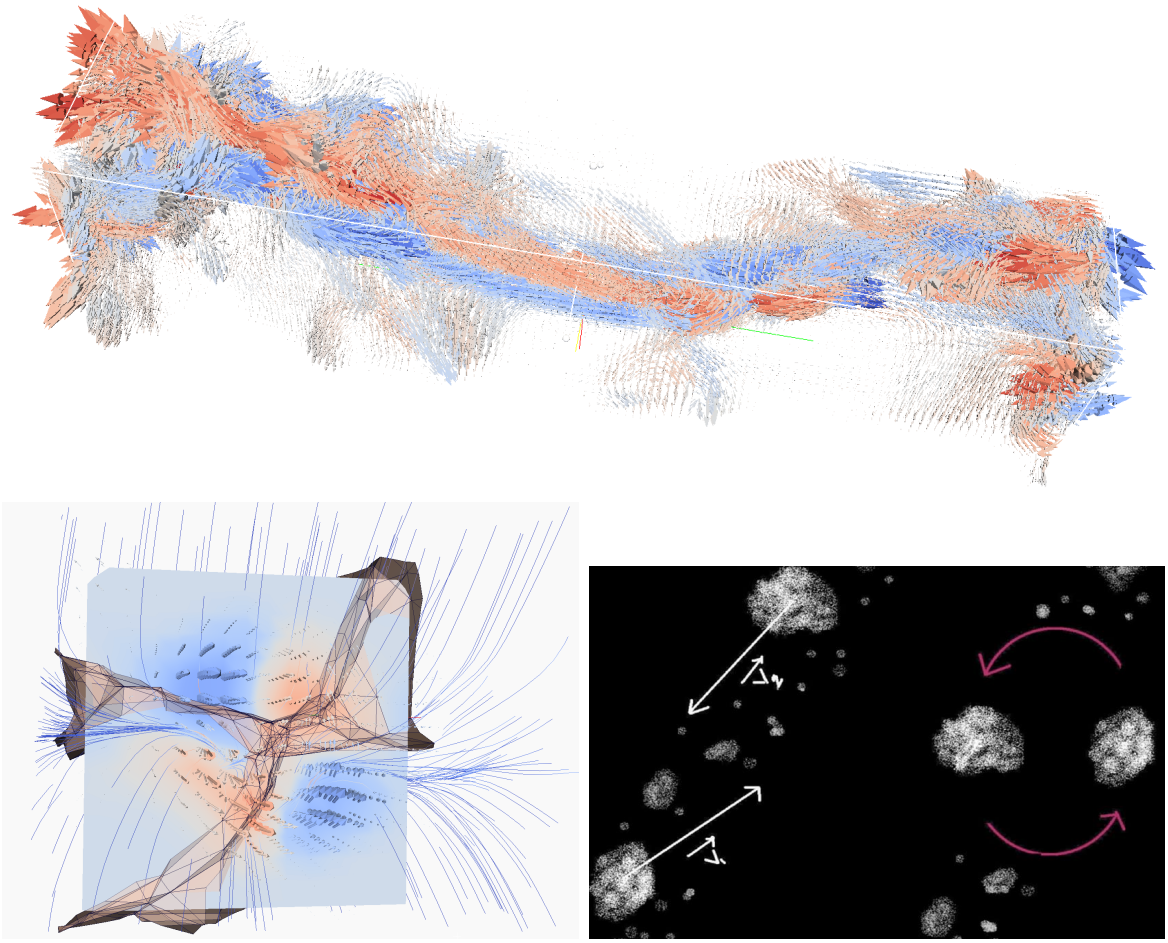


Figure B.3: En haut: champ de vorticité le long du filament. La couleur dépend de la composante algébrique le long du filament (flèches rouges en allant vers la gauche du filament, bleues en allant vers la droite). En bas à gauche: coupe typique perpendiculaire au filament. En bas à droite: schémas de la fusion de deux halos le long du filament. Le moment angulaire orbitale avant fusion est converti en spin pour l'objet résultant.

filaments: dans les murs, le boost en densité n'est pas suffisant pour permettre leur formation.

Il faut noter que les noeuds de la toile ne sont pas forcément le lieu de naissance des galaxies mais la destination vers laquelle elles migrent. Elles vont donc en partie garder en mémoire l'anisotropie du lieu qui les a vu naître, par exemple par le biais de leur spin. Pendant la migration, les galaxies coalescent et peuvent ainsi perdre une partie de leurs propriétés initiales.

Dans ce contexte, nous avons proposé de revisiter la théorie des champs de marée en tenant compte de l'anisotropie de l'environnement filamentaire. Cette théorie anisotrope est totalement analytique et se base sur la théorie des champs aléatoires contraints.

En régime linéaire, le spin généré par les champs de marée peut s'écrire

$$L_i \propto \epsilon_{ijk} I_{jl} T_{lk} \,, \tag{B.2}$$

où T_{ij} est le tenseur de marée soit le Hessien détracé du potentiel gravitationnel et I_{ij} est le tenseur d'inertie de la proto-galaxie. Afin de construire une théorie locale, il est nécessaire d'adopter une prescription locale pour définir le tenseur d'inertie (objet intrinsèquement non-local). Ici, nous avons fait l'hypothèse que le Hessien du champ de densité est une bonne approximation, ce que l'on peut justifier par le fait que les proto-galaxies ont tendance à se former aux pics du champ de densité et devraient donc aussi hériter de leur forme qui est donnée par le tenseur des dérivées secondes de la densité. Nous choisissons donc de définir le spin comme

$$s_i \equiv \epsilon_{ijk} H_{jl} T_{lk} \,. \tag{B.3}$$

L'avantage de cette prescription est qu'elle ne dépend que des dérivées secondes (tenseur de marée) et quatrième (Hessien de la densité) du potentiel. Nous supposons ici que tous les champs sont de variance unité. Dans les conditions initiales supposées Gaussiennes, la PDF jointe du tenseur de marée et du Hessien de la densité (que nous rassemblons dans un vecteur $\mathbf{X} = (\{T_{ij}\}_{i\leq j}, \{H_{ij}\}_{i\leq j})$) s'écrit

$$\mathcal{P}(\mathbf{X}) = \frac{\exp\left(-\frac{1}{2}\mathbf{X}^{\mathrm{T}} \cdot \mathbf{C_0}^{-1} \cdot \mathbf{X}\right)}{\det|\mathbf{C_0}|^{1/2}\,(2\pi)^{(\mathrm{d}+1)(\mathrm{d}+2)/2}} \,, \tag{B.4}$$

où $\mathbf{C} \equiv \langle \mathbf{X} \cdot \mathbf{X}^{\mathrm{T}} \rangle$ est la matrice des variances-covariances qui peut être facilement calculée. Par exemple, à 2D, si $\mathbf{X} = (T_{11}, T_{12}, T_{22}, H_{11}, H_{12}, H_{22})$ alors

$$\mathbf{C} = \begin{pmatrix} 3/8 & 0 & 1/8 & -3\gamma/8 & 0 & -\gamma/8 \\ 0 & 1/8 & 0 & 0 & -\gamma/8 & 0 \\ 1/8 & 0 & 3/8 & -\gamma/8 & 0 & -3\gamma/8 \\ -3\gamma/8 & 0 & -\gamma/8 & 3/8 & 0 & 1/8 \\ 0 & -\gamma/8 & 0 & 0 & 1/8 & 0 \\ -\gamma/8 & 0 & -3\gamma/8 & 1/8 & 0 & 3/8 \end{pmatrix} ,$$

avec $\gamma = \frac{\sigma_1^2}{\sigma_0 \sigma_2}$ $(= \sqrt{(n+2)/(n+4)}$ pour un spectre de puissance du champ de densité $P(k)$ en loi de puissance d'indice n et les variances sont données par

$$\sigma_n^2 = \int_0^\infty \int_0^{2\pi} \mathrm{d}\theta \; \mathrm{d}k \; P(k) k^{2n+1} \,.$$

Il est ainsi aisé de vérifier que l'espérance du spin est nulle $\langle s_i \rangle = \epsilon_{ijk} \langle H_{jl} \rangle \langle T_{lk} \rangle = 0$ car tous les termes qui composent cette contraction antisymétrique sont des produits de H_{jl} et T_{lk} qui sont décorrélés (comme par exemple $H_{12}T_{23}$). C'est un résultat attendu car il n'y a pas de direction privilégiée dans le problème.

Cependant les galaxies ne se forment pas partout mais préférentiellement dans les filaments et les noeuds qui trouvent leur origine dans les assymétries du champ aléatoire initial décrivant l'Univers primordial ensuite amplifiées par l'effondrement gravitationnel. La présence de ces structures induit des directions privilégiées pour le tenseur de marée ainsi que le tenseur d'inertie et donc en fin de compte pour le spin. Il est donc naturel de se demander quel est le spin moyen au voisinage d'un filament typique. Puisqu'un filament peut être défini comme la ligne de champ joignant deux maxima de la densité en passant par un point selle (gradient nul et deux courbures négatives), le problème se réduit à l'étude de l'espérance du spin au voisinage d'un point selle de type filament ayant une géométrie donnée (qui impose la direction du filament et du mur). Ce spin contraint $\langle s_i(\mathbf{r})|\text{saddle}\rangle$ dépend de $\mathbf{r}$, la séparatrice entre le proto-objet et le point selle (ou pic à 2D si on considère des coupes perpendiculaires au filament). Cette quantité est facilement déductible de la PDF jointe des dérivées secondes et quatrièmes du potentiel en deux points séparés par $\mathbf{r}$. Nous devons aussi introduire le gradient de la densité (qui est une dérivée troisième du potentiel) $\mathbf{q} = \nabla\delta/\sigma_1$ afin d'imposer une condition de point critique $\delta_D(x-\nu)|\det H_{ij}|\delta_D(q_i)$. La géométrie du point critique peut aussi être imposée par la donnée des valeurs propres du Hessien de la densité λ_i. Si nous rassemblons le tenseur de marée, le gradient et le Hessien de la densité dans un vecteur $\mathbf{X}$ à l'emplacement du proto-objet et $\mathbf{Y}$ à l'emplacement du point critique, cette PDF à deux point s'écrit

$$\mathcal{P}(\mathbf{X},\mathbf{Y}) = \frac{1}{\sqrt{\det|2\pi\mathbf{C}|}} \exp\left(-\frac{1}{2}\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}^{\mathrm{T}} \cdot \mathbf{C}^{-1} \begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}\right), \tag{B.5}$$

où $\mathbf{C}_0 \equiv \langle \mathbf{X}\cdot\mathbf{X}^{\mathrm{T}}\rangle$, $\mathbf{C}_\gamma \equiv \langle \mathbf{X}\cdot\mathbf{Y}^{\mathrm{T}}\rangle$ et

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_0 & \mathbf{C}_\gamma\\ \mathbf{C}_\gamma^{\mathrm{T}} & \mathbf{C}_0\end{bmatrix}.$$

Encore une fois, ces matrices peuvent être calculées à partir du spectre de puissance $P(k)$.

Finalement, l'espérance du spin, au voisinage d'un point selle de type filament ayant une géométrie donnée $\langle s_i(\mathbf{r})|\text{saddle}\rangle$, peut être évaluée analytiquement à partir de la PDF jointe. Ce résultat est donné dans [Codis et al., 2015b]. En résumé, dans le plan du point selle, les spins sont alignés en moyenne avec l'axe du filament et forment quatre quadrants de parité opposée. En se rapprochant des noeuds, les spins deviennent de plus en plus perpendiculaires au filament. Cette géométrie est illustrée sur la figure B.4.

L'extension anisotrope de la théorie des champs de marée présentée ici prédit donc une géométrie quadrupolaire dans les filaments en accord avec les quadrants de vorticité obervés dans les simulations ([Laigle et al., 2015]). Il est frappant de constater que la théorie Lagrangienne décrite ici capture les processus Eulériens a priori non-linéaires de bascule de spin par fusions. En plus des quatre quadrants, cette théorie prédit une transition spatiale de la direction du spin: alignée avec le filament proche du point selle et perpendiculaire au filament vers les noeuds. Cette transition dans l'espace peut être traduite en une transition en masse (faibles masses ayant un spin aligné et objets massifs un spin perpendiculaire) si l'on réalise que la distribution locale des masses des halos varie le long du filament (faibles masses loin du filament et fortes masses vers les

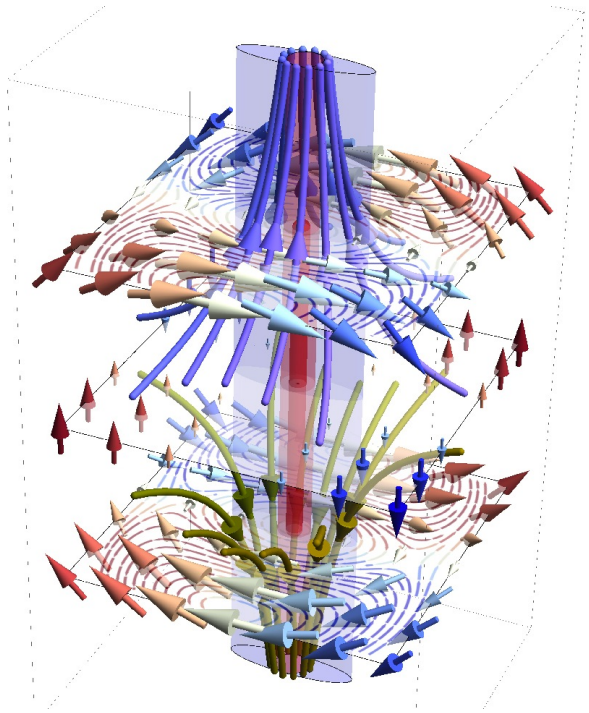


Figure B.4: Flots de spin et de vitesse au voisinage d'un filament vertical (en rouge) au sein d'un mur (mauve). Les flèches rouges et bleues indiquent la distribution de spin alors que les trois coupes horizontales montrent les lignes de champs du spin projeté dans le plan correspondant.

noeuds). Ceci est dû encore une fois au boost en densité généré par la présence du filament. Le formalisme de croisement de barrières ([Peacock and Heavens, 1990, Bond et al., 1991, Paranjape et al., 2012, Musso and Sheth, 2012]) permet alors d'associer une masse typique pour des objets se formant dans des régions de densité à grande échelle donnée. Dans ce contexte, la théorie ATTT prédit une orientation du spin qui, en moyenne, dépend de la masse avec une masse de transition $\mathcal{M}_{\rm tr}$ pour la bascule du spin ($\langle \cos\theta \rangle = 0.5$) d'environ $4 \cdot 10^{12} M_\odot$, si on choisit de définir le filament avec une échelle de lissage de 5 Mpc/h, ce qui est en accord avec la mesure faite dans [Codis et al., 2012].

B.2.4 Vers une quantification des effets d'alignements intrinsèques

La cohérence à grande échelle de certaines propriétés des galaxies telles que le spin peut être un contaminant important des observables de lentillage faible. Etant donné d'une part la nature anisotrope des grandes structures et son impact complexe sur les formes et spins des galaxies et d'autre part la dépendance forte avec les propriétés physiques des galaxies, il est probablement difficile de se reposer uniquement sur le théorie linéaire isotrope (par exemple [Lee and Pen, 2001]) ou les simulations de matière noire seule pour prédire et contrôler les alignements intrinsèques. Avec l'avènement des simulations hydrodynamiques cosmologiques, nous sommes maintenant en position pour mesurer directement cet effet dans ces simulations plutôt que d'utiliser le modèle du halo ou des modèles semi-analytiques ([Schneider and Bridle, 2010, Joachimi et al., 2013b]). Dans [Codis et al., 2015a]), nous utilisons la simulation Horizon-AGN présentée dans [Dubois et al., 2014] à redshift $z = 1.2$ pour mesurer le niveau d'alignements intrinsèques lorsque l'on prend le spin comme proxy pour la forme des galax-

ies. Si cette hypothèse permet d'utiliser une quantité – le spin – plus robuste face aux effets de résolution numérique que le tenseur d'inertie, il faut noter que cette approximation n'est probablement valable que pour les galaxies de type disque dominées par leur rotation.

C'est dans ce contexte que nous avons donc mesuré dans un premier temps la corrélation en fonction de la séparation entre le spin des galaxies d'une part et le tenseur d'inertie d'autre part. A séparation nulle, les résultats de [Codis et al., 2012, Dubois et al., 2014] sont retrouvés à savoir que les galaxies ont un spin corrélé avec la direction du filament (qui est bien alignée avec l'axe mineur du tenseur d'inertie) de façon masse et couleur dépendant. Au-delà de la statistique à un point, il est intéressant d'étudier comment ce signal varie avec la séparation. Parce que le tenseur de marée au voisinage de la galaxie contribue aussi au signal de lentillage faible porté par des galaxies plus lointaines, il est clair que les corrélations croisées spin – tenseur de marée sont étroitement liées au terme GI. Comme on peut l'espérer, l'axe du spin et les directions propres du tenseur de marée se décorrèlent à mesure que la séparation augmente. Cependant, alors que les corrélations spin – axe intermédiaire du tenseur de marée deviennent négligeables sur des échelles $r > 3\ h^{-1}$ Mpc, le signal persiste sur des distances allant jusqu'à $\sim 10\ h^{-1}$ Mpc pour les axes mineur et majeur du tenseur de marée (ce qui n'est rien d'autre que la taille typique d'un filament). Ces corrélations à grande portée représentent potentiellement une source d'inquiétude pour les expériences basées sur la mesure de l'astigmatisme cosmique qui pourraient donc être contaminées par une dose non-négligeable d'alignements intrinsèques de type "GI". Une compréhension plus fine du taux de contamination requiert une modélisation des ellipticités projetées de galaxies dans la simulation ainsi qu'une modélisation des vraies observables en tenant compte des systématiques observationnels telles que l'atténuation par la poussière des couleurs de galaxies. Il est toutefois amusant de noter que le comportement de la fonction de corrélation à deux points entre spin et tenseur de marée peut être comprise théoriquement par le simple usage d'un champ aléatoire Gaussien δ pour lequel on calcule la PDF jointe des dérivées secondes du potentiel associé (ϕ_{ij}, ϕ étant relié à δ par l'équation de Poisson). A partir de cette PDF, il est trivial de calculer l'angle moyen entre les directions propres de ϕ_{ij} calculé en deux localisations séparées de r. Une fois que cette fonction est multipliée par un facteur ad hoc permettant de retrouver la valeur de la statistique à un point, on trouve (voir figure B.5, panneau de droite) le même comportement avec la distance que ce qui est mesuré dans la simulation (voir figure B.5, panneau de droite).

Une autre quantité intéressante – cette fois reliée au terme II – est la mesure des auto-corrélations d'ellipticités par le biais de la fonction de corrélation à deux points du spin. Dans [Codis et al., 2015a], il apparaît que les axes de rotation des galaxies bleues sont significativement corrélés sur des distances de l'ordre de 10 Mpc/h mais que les galaxies rouges ne présentent aucune corrélation notable. Afin de se rapprocher des observables qui utilisent l'astigmatisme cosmique, la question qui se pose est de savoir quelle fraction du signal perdure après projection sur le ciel. Pour cela, nous avons utilisé une approximation de disque mince et montré que les ellipticités projetées ainsi obtenues étaient corrélées sur des distances angulaires de l'ordre de 10 arcminutes pour les galaxies bleues alors qu' elles ne présentent pas de corrélations pour les rouges, en adéquation avec l'étude à 3 dimensions. Il est à noter que ce résultat n'est pas en contradiction avec les observations car il est à plus haut redshift ($z = 1.2$). Afin d'améliorer ces premiers résultats, nous nous sommes intéréssés à l'évolution de ces alignements avec le temps et avons amélioré notre prescription pour les ellipticités des galaxies en utilisant le tenseur d'inertie plutôt que le spin. La figure B.6 montre ainsi la moyenne du cosinus carré de l'angle entre les axes mineurs du tenseur d'inertie en fonction de la séparation et pour différentes classes de masse. Un alignement est clairement détecté pour toutes les masses et pour des distances allant jusqu'à $\approx 10\ h^{-1}$ Mpc. Ce signal est bien entendu similaire aux résultats obtenus avec les spins puisque spin et axe mineur du tenseur d'inertie sont des quantités fortement corrélées.

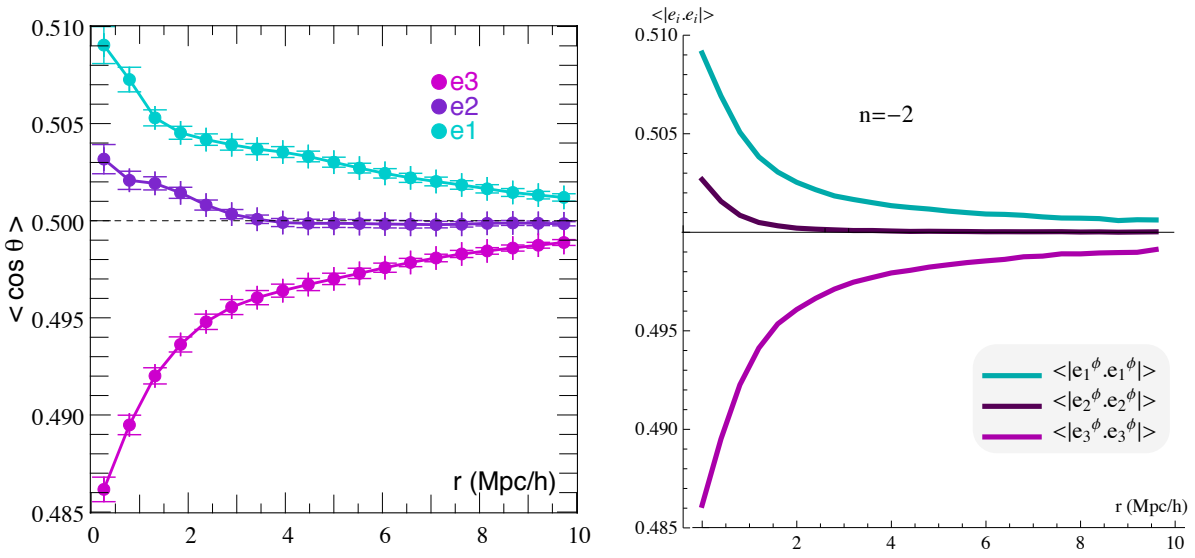


Figure B.5: A gauche: Cosinus moyen de l'angle entre le spin des galaxies et l'axe mineur (cyan), intermédiaire (violet) et majeur (magenta) du tenseur de marée en fonction de la séparation. A droite: Idem pour un champ Gaussien aléatoire de spectre de puissance en loi de puissance à un facteur ad hoc près.

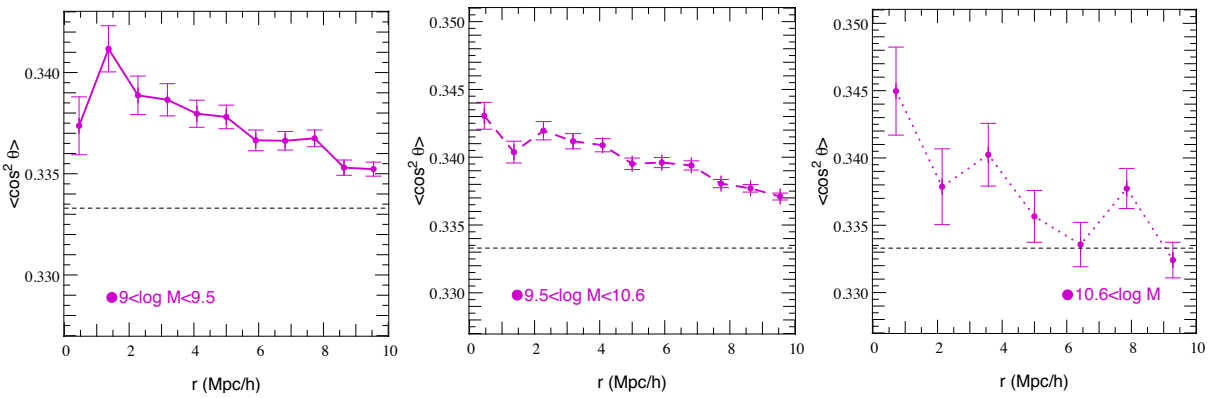


Figure B.6: Moyenne du cosinus carré de l'angle entre les axes mineurs du tenseur d'inertie en fonction de la séparation et pour différentes classes de masse. Cette mesure est faite à redshift $z = 0.5$. La différence principale se situe au niveau des objets les plus massifs qui sont principalement des elliptiques et pour lesquels le spin – qui est plutôt mal défini – et l'axe mineur sont moins bien alignés.

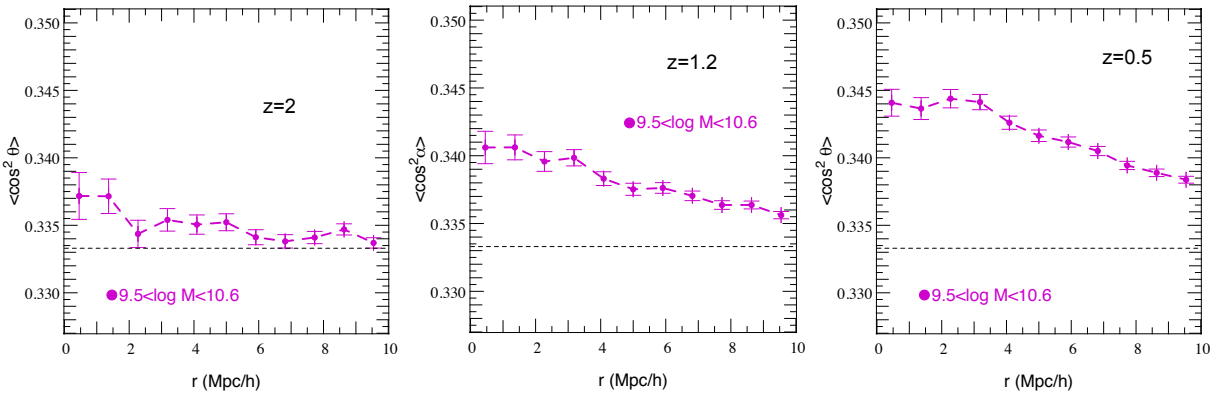


Figure B.7: Corrélation spin-spin en fonction de la séparation pour des masses stellaires: $3 \times 10^9 < M_\mathrm{s} < 4 \times 10^{10}\,\mathrm{M}_\odot$ à redshift $z = 2$ (à gauche), $z = 1.2$ (en haut au centre), et $z = 0.5$ (à droite). En cas de répartition uniforme, l'espérance est de 1/3 (représentée avec des tirets ici).

L'évolution en redshift est montrée sur la figure B.7 pour trois redshifts: $z = 2$, 1.2 and 0.5. Les corrélations tendent à croître avec le temps. Des résultats similaires sont obtenus quand le tenseur d'inertie est utilisé.

L'inquiétude principale dans ce type d'étude reste les effets d'alignement avec la grille qui sont propres aux simulations AMR. En effet, comme les forces sont calculées sur une grille, les spins et les tenseurs d'inertie ont tendance à s'aligner avec les directions de la grille. Cet effet génère donc des corrélations additionnelles qui sont difficiles à séparer du signal physique que l'on cherche à mesurer. Différentes stratégies peuvent être proposées pour tenter d'évaluer ou de minimiser cet effet. Une simple redistribution aléatoire des galaxies ne peut pas suffire car l'effet d'alignement avec la grille est probablement corrélé à petite échelle: deux galaxies proches ressentent un champ de marée similaire et ont donc tendance à s'aligner similairement avec la grille. Une approche alternative pourrait être de sélectionner un échantillon (en terme de masse, couleur, âge stellaire, etc) de galaxies parmi les moins contaminées mais cette procédure réduit considérablement le nombre de galaxies et donc la statistique. De surcroît, il est difficile de voir comment ces sélections affectent la mesure d'alignement intrinsèque et sa comparaison aux données observationnelles. Cependant, il est possible d'étudier une autre observable – que l'on utilise beaucoup quand il s'agit d'alignements intrinsèques – la fonction de corrélation orientation-séparation

$$\eta_r(r) = \langle |\hat{\mathbf{r}} \cdot \hat{\mathbf{e}}(\mathbf{x}+\mathbf{r})|^2 \rangle - 1/3\,, \tag{B.6}$$

où $\hat{\mathbf{e}}$ est le vecteur unité le long de l'axe mineur du tenseur d'inertie et $\hat{\mathbf{r}}$ le vecteur séparation. Une corrélation positive indique une tendance pour la séparation et l'axe mineur à être parallèle ce qui signifie que la galaxie est étirée tangentiellement par rapport à la seconde. Une corrélation négative indique que la galaxie est orientée radialement par rapport à la seconde. L'avantage de cette observable est qu'elle est moins sensible aux effets d'alignement avec la grille car la distribution spatiale des galaxies – et donc l'ensemble des vecteurs séparation – n'a pas de raison d'y être sujette (comme nous le montrons dans [Chisari et al., 2015]). Dans cet article, nous avons utilisé la simulation Horizon-AGN et montré que les elliptiques ont tendance à s'aligner radialement par rapport aux sur-densités de la matière noire et par rapport aux autres elliptiques. Cette tendance est en accord avec les observations mais l'amplitude dépend fortement de la façon de mesurer les formes et de l'échantillon de galaxies utilisé dans la simulation. Les galaxies spirales ont, elles, tendance à être orientées tangentiellement autour des elliptiques. Si ce signal est supprimé en projection, nous ne pouvons pas garantir que les alignements de disques peuvent être ignorés pour les futures expériences utilisant l'astigmatisme cosmique.

Ce travail autour de la mesure des alignements intrinsèques à l'aide de simulations hydrodynamiques est un premier pas vers une modélisation fine de cet effet et ouvre la voie à des études plus réalistes. Une amélioration importante serait par exemple de mesurer les ellipticités à l'aide de la luminosité et non de la répartition de masse, de faire des mesures sur le cône de lumière et de tenir compte de différentes systématiques observationnelles.

B.2.5 Conclusion

Nous avons présenté ici différentes investigations qui ont conduit notamment à la publication de quatre articles ([Codis et al., 2012, Codis et al., 2015b, Codis et al., 2015a, Chisari et al., 2015]). En résumé, nous avons montré que les halos et les galaxies sont fortement corrélés avec la toile cosmique en mettant une emphase particulière sur les spins et les ellipticités car ce sont des quantités importantes en ce qui concerne notre compréhension de la formation des galaxies et la prise en compte des effets d'alignements intrinsèques. Ces corrélations – qui dépendent notamment du temps et de la masse – peuvent être comprises analytiquement par le biais d'une théorie anisotrope des effets de marée (ATTT). C'est évidemment un grand succés de notre paradigme de formation des galaxies mais il faut aussi garder en tête que ces corrélations sont générées par la dynamique (linéaire!) à grande échelle. Les non-linéarités à petite échelle diluent cette cohérence, de sorte que l'on ne peut pas croire que la théorie linéaire des champs de marée puisse prédire un à un le spin des galaxies à bas redshift et cela a en effet déjà été souligné dans différentes études numériques. L'histoire individuelle de chaque galaxie comprenant fusions et accrétion de matière contribue de façon largement non-linéaire à l'évolution récente des spins qui vont donc naturellement dévier de la prédiction de la théorie linéaire.

Les alignements intrinsèques sont donc élégamment compris comme le résultat des flots de matière à grande échelle. Cependant, si l'on souhaite les prédire précisément, il faut clairement aller au-delà de la théorie linéaire ou des modèles semi-analytiques qui peuvent simplement donner une image globale des processus en jeu. Les simulations hydrodynamiques sont le seul moyen d'évaluer la contamination par les alignements intrinsèques que l'on peut attendre pour les missions futures basées sur l'utilisation de l'astigmatisme cosmique. En effet, celle-ci semble dépendre fortement de la dynamique non-linéaire du gaz, de la physique baryonique et donc des effets de sélection sur la base par exemple de la couleur, la masse ou la morphologie des galaxies mais aussi des systématiques observationnelles. Seules les simulations numériques hydrodynamiques peuvent tenir compte efficacement de l'ensemble de ces effets. La contrepartie est que ces simulations ont beaucoup de paramètres à ajuster qui encodent notre ignorance sur une partie de la physique sous-grille, ce qui pose diverses questions: à quel point nos prédictions dépendent-elles du code et des détails de la physique sous-grille utilisée? Quels sont les effets de l'utilisation d'une méthode AMR par rapport à du SPH? A quel point est-ce que la rétroaction des noyaux actifs de galaxies jouent un rôle en alignant ou désalignant les galaxies? Toutes ces questions pour le moment sans réponse représentent un formidable terrain d'investigation et devraient être cruciales dans les années à venir.

B.3 Les grandes structures comme sonde cosmologique

La cosmologie d'aujourd'hui fait la part belle aux champs aléatoires qui sont des ingrédients fondamentaux de notre description du champ de matière à grande échelle et du fond diffus cosmologique. La structuration à grande échelle de la distribution de matière dans l'Univers

semble être le résultat de la croissance de petites fluctuations primordiales quasi-Gaussiennes sous les effets de la Gravitation. L'apparition de non-linéarités de plus en plus fortes est due à la dynamique non-linéaire des structures en formation mais une faible non-Gaussianité peut aussi être présente initialement dans les inhomogénéités de l'Univers primordial. C'est pour cette raison que l'étude de signatures non-Gaussiennes dans les données observationnelles permet d'étudier les détails de la physique à l'oeuvre dans l'Univers primordial et le mécanisme de croissance des structures avec ainsi la possibilité de lever le voile sur différentes inconnues telles que le contenu de l'Univers, le rôle du biais entre galaxies et matière noire et les raisons de l'accélération de l'expansion de l'Univers (énergie noire? modifications de la gravité d'Einstein?).

Avec l'arrivée de grands relevés de galaxies, les astronomes entrent dans l'ère de la cosmologie statistique et voient arriver une quantité impressionnante de données. La question se pose donc de construire des outils qui permettront d'extraire le plus d'information possible de ces données, ce qui signifie en particulier de sonder le régime non-linéaire de formation des structures. La façon la plus commune de procéder se base sur l'utilisation des fonctions de corrélations à N points qui caractérisent la distribution observée de galaxies. Dans l'Univers primordial Gaussien, le champ de densité de matière est complètement décrit par la fonction de corrélation à deux points. Des départs à la Gaussianité apparaissent ensuite dès que la dynamique devient non-linéaire soit à petites échelles ou temps longs et ajoutent de l'information qui n'est plus uniquement contenue dans la fonction à deux points. Cependant, les fonctions de corrélations à N points sont de plus en plus difficiles à mesurer. Elles sont bruitées, sujettes à la variance cosmique et sensibles aux systématiques telles que la géométrie complexe du relevé. D'un point de vue théorique, elles sont également de plus en plus difficiles à prédire dans le contexte de la théorie des perturbations cosmologiques. Typiquement, le spectre de puissance à l'ordre 2, le bispectre à l'ordre 1 et le trispectre à l'ordre 0 sont accessibles avec les méthodes standard mais il est bien difficile d'aller plus loin à la fois dans le régime non-linéaire (et donc dans les ordres du développement perturbatif) et dans le nombre de champs (N) même si des schémas de resommations ont été mis en place et permettent d'atteindre des précisions de l'ordre du pourcent dans le régime faiblement non-linéaire (voir par exemple [Taruya et al., 2012] pour le spectre de puissance à l'ordre 2 de la théorie des perturbations régularisée RegPT).

Il est donc essentiel de trouver des estimateurs alternatifs pour extraire de l'information du régime non-linéaire de formation des structures afin de compléter les outils traditionnels. C'est en particulier critique si l'on souhaite en apprendre davantage sur l'origine de l'énergie noire qui représente près de 70 % du budget énergetique de notre Univers! Nous présentons ici deux types d'observables cosmologiques qui semblent prometteuses: des estimateurs topologiques et géométriques (comme les fonctionnelles de Minkowski ou les comptages d'extrema) et la statistique de comptage dans des cellules concentriques.

B.3.1 Caractérisation topologique et géométrique des grandes structures

La première des deux observables que nous présentons ici utilisent les caractéristiques topologiques et géométriques du champ de densité. La topologie d'un champ de dimension 3 peut être entièrement décrite par seulement 4 fonctionnelles qui portent le nom de Minkowski et qui peuvent être calculées analytiquement quand les non-Gaussianités sont suffisamment faibles. En plus de ces 4 fonctionnelles de Minkowski, on peut aussi s'intéresser à des observables similaires qui, elles, décrivent la géométrie de la toile cosmique: comptage de pics, de vides, propriétés du squelette telles que sa longueur, etc. L'avantage de tous ces estimateurs par rapport aux fonctions à N points est double. D'une part, ils dépendent faiblement de tout biais monotone.

Par exemple, le nombre de maxima d'un champ et de toute transformée monotone de ce champ est le même. D'autre part, la topologie du champ est, par essence, une caractéristique robuste et donc typiquement plus propre à mesurer car moins sensible aux masques et bruits divers.

Dans [Codis et al., 2013], nous avons fait la théorie de ces estimateurs pour des champs faiblement non-Gaussiens en espace des redshifts. Le résultat dépend des moments du champ et de ses dérivées qui sont eux-mêmes prédits par la théorie perturbative en fonction de la variance σ^2 et du paramètre de distorsion en redshift β. Avec les prochains relevés spectroscopiques 3D comme Euclid, l'analyse statistique de la topologie de l'Univers décalé vers le rouge va permettre de mesurer de façon robuste certaines combinaisons de moments en fonction du redshift et donc de quantifier l'évolution de l'équation d'état de l'énergie noire au cours du temps et de possibles modifications de la gravité d'Einstein. Dans [Codis et al., 2013], nous avons spécifiquement étudié comment les fonctionnelles de Minkowski en espace des redshifts pourraient permettre d'estimer l'évolution cosmique de la croissance des structures par le biais du paramètre β.

Des ensembles critiques non-Gaussiens aux cumulants Considérons les propriétés statistiques des ensembles critiques d'un champ aléatoire (la densité par exemple) telles que la caractéristique d'Euler des contours de densité au dessus d'un seuil, le nombre de maxima de contraste donné ou encore la longueur de squelette au dessus d'un seuil en densité. Le calcul de leur espérance statistique requiert la donnée de la densité de probabilité jointe du champ et de ses dérivées premières et secondes. Celle-ci peut être calculée facilement en fonction du spectre de puissance si la statistique est Gaussienne. Si le champ est non-Gaussien, elle peut être développée autour d'une Gaussienne sous la forme d'une série dite de Gram-Charlier.

Une fois cette PDF connue, tous les estimateurs mentionnés dans cette section peuvent être calculés en la marginalisant sous certaines conditions. Par exemple, pour calculer la densité de points critiques, il faudra imposer une condition de gradient nul. Après quelques pages d'algèbre, les prédictions sont obtenues à tous les ordres du développement en non-Gaussianité et en fonction du contraste (par le biais typiquement de polynômes de Hermite) et des cumulants du champ sous-jacent (qui apparaissent dans les coefficients de Gram-Charlier).

Par exemple, $\mathcal{N}_3(\nu)$, l'aire (par unité de volume) des isosurfaces de densité $x=\nu$ en espace des redshifts s'écrit

$$\mathcal{N}_3 = \frac{2e^{-\nu^2/2}\sigma_1}{\sqrt{3}\pi\sigma}\left[1+\frac{\langle x^3\rangle}{3!}H_3(\nu)+\left(\frac{1}{3}\langle xq_\perp^2\rangle\left[1-\frac{4}{15}\beta_\sigma\right]+\frac{1}{6}\langle xx_3^2\rangle\left[1+\frac{8}{15}\beta_\sigma\right]\right)H_1(\nu)\right]+\mathcal{O}(\beta_\sigma^2)\,,$$

où β_σ est un paramètre d'anisotropie qui est uniquement fonction de β et $(\mathbf{q}_\perp, x_3)$ est le gradient du champ x.

Comparaison avec des simulations numériques A suffisamment grand lissage, les prédictions sont en accord avec les mesures faites dans des simulations numériques à N corps. Quand le lissage diminue, la prédiction au premier ordre du développement commence à devier de la mesure dans les ailes de la distribution. L'effet de la distorsion en redshift sur les estimateurs 3D comme la caractéristique d'Euler semble faible dans la simulation Horizon 4π (voir figure B.8). Cela signifie que la théorie isotrope est en fait suffisante et peut donc permettre de mesurer la variance des fluctuations de densité en utilisant la première correction non-Gaussienne comme montré par [Gay et al., 2012]. Cela signifie aussi qu'il y a peu d'espoir d'extraire une information cosmologique à partir des distorsions de l'espace des redshifts en utilisant cette méthode.

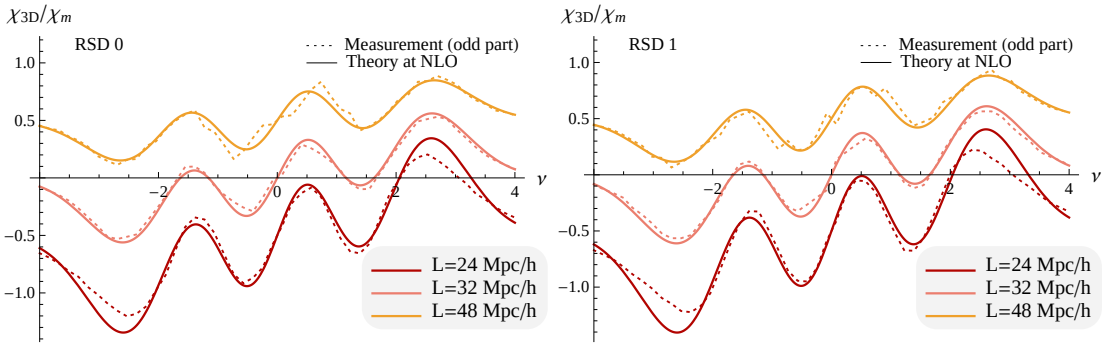


Figure B.8: *A gauche*: première correction non-Gaussienne de la caractéristique d'Euler à 3D en fonction du lissage telle que mesurée (en pointillés) et prédite (trait plein) dans la simulation Horizon 4π (espace réel). Chaque courbe est normalisée par le maximum de la composante Gaussienne et est décalée par lissage à des fins de clarté. *A droite*: Même résultat mais cette fois en espace des redshifts. Pour les lissages utilisés ici, la théorie au premier ordre est en bonne accord avec la simulation, en particulier pour des contrastes faibles et intermédiaires. La différence produite par l'effet de distorsion redshift est faible.

Cependant, il apparaît que les caractéristiques topologiques 1 et 2D (sur des lignes et des plans) sont elles affectées par les distorsions en redshift (voir figure B.9) et cet effet prédit par la théorie est en accord avec les simulations. Ceci devrait donc permettre de mesurer β et donc potentiellement d'obtenir plus d'information sur notre Univers.

Des cumulants à la cosmologie Dans le régime où la théorie des perturbations est valide en espace des redshifts, tous les cumulants qui entrent dans l'expression des fonctionnelles de Minkowski peuvent être prédits en fonction du modèle cosmologique adopté. Ainsi, la mesure dans les données de ces fonctionnelles permet en principe de mettre des contraintes sur la cosmologie en utilisant certaines combinaisons de cumulants. Une analyse en redshift permet a priori d'obtenir plus d'information qu'une analyse en espace réel. En effet, les cumulants qui impliquent la ligne de visée et ceux qui impliquent les directions transverses (le plan du ciel) sont qualitativement différents. Cette différence encode une information dynamique (sur les vitesses). En principe, la partie anisotrope de ces cumulants peut donc être utilisée pour tester notre théorie de la gravité dans le contexte de la théorie perturbative de croissance des grandes structures. Les statistiques topologiques et géométriques 3D ne permettent pas en elles-mêmes de déterminer séparément les cumulants impliquant la ligne de visée de ceux impliquant les directions du ciel. Pour mesurer la contribution anisotrope, il est préférable de s'attacher à l'étude de tranches du volume 3D orientées différemment par rapport à la ligne de visée (avec un angle $\theta_{\mathcal{S}}$). Par exemple, en mesurant le longueur des isocontours dans ces tranches, $\mathcal{N}_2$, on obtient une mesure séparée de $\langle xq_\perp^2 \rangle$ et $\langle xx_3^2 \rangle$.

Une expérience de pensée cosmologique Comme on vient de le mentionner, la dépendance angulaire des fonctionnelles de Minkowski 2D permet en principe de mesurer β. Au-delà d'un simple effet sur l'amplitude globale (qui est suffisant pour mesurer β seul), cette dépendence apparaît également dans la première correction non-Gaussienne. Si la valeur de β est importante en elle-même (pour exemple pour étudier le biais et tester de possibles modifications de la gravité d'Einstein), il est aussi important pour mesurer $D(z)$ le taux de croissance des structures car il

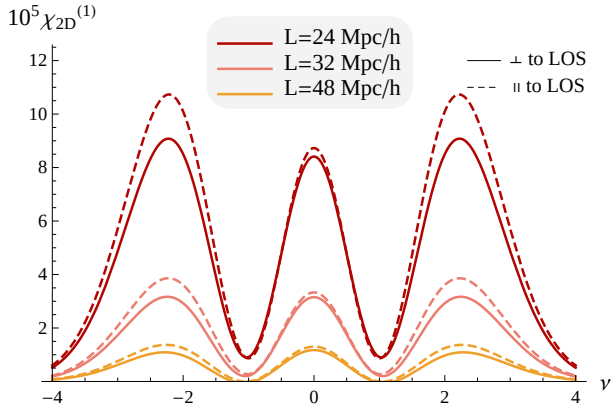


Figure B.9: Prédiction de la caractéristique d'Euler 2D dans l'espace des redshifts le long (tirets) et perpendiculaire (trait plein) à la ligne de visée avec les cumulants mesurés dans la simulation Horizon 4π . La dépendance avec l'angle entre la tranche et la ligne de visée est claire, en particulier pour les faibles lissages.

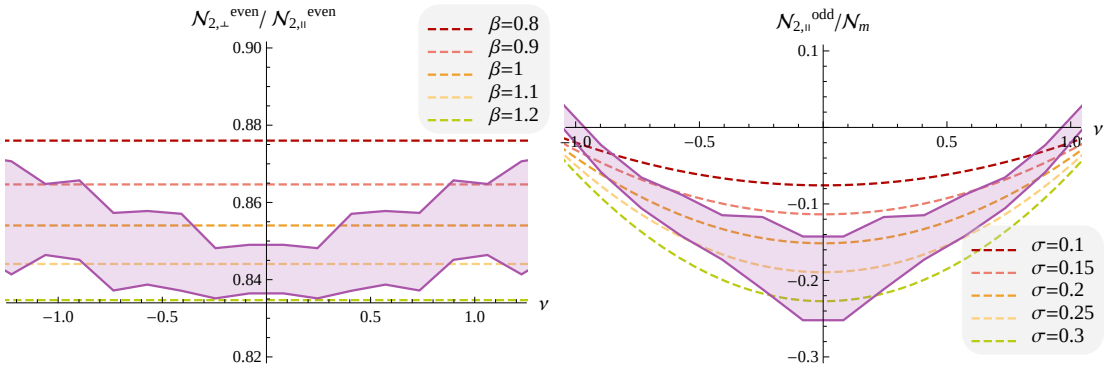


Figure B.10: Reconstruction de β à partir de la partie Gaussienne de $\mathcal{N}_2$ (à gauche) et σ à partir de la première correction non-Gaussienne (à droite).

permet de relier la dispersion en espace des redshifts, σ^2, à sa valeur en espace réel, σ_0^2 qui est elle-même proportionnelle à $D(z)$. Une façon de procéder est donc d'utiliser le terme Gaussien pour poser des contraintes sur β puis le premier terme non-Gaussien pour σ (dont l'amplitude n'est pas contrainte par la partie Gaussienne seule). Nous avons appliqué ce schéma pour mesurer β et σ_0 dans une simulation. Les fonctionnelles de Minkowski sont mesurées dans un ensemble de 19 simulations invariantes d'échelle (n=-1) de matière noire (256^3 lissées sur 15 pixels ce qui correspond à $\sigma_0 = 0.18$). En mesurant la partie paire (dominée par la composante Gaussienne) de $\mathcal{N}_2$, on obtient les contraintes suivantes sur β: $\hat{\beta} = 1.04 \pm 0.05$, illustrées sur la figure B.10 (à gauche). En utilisant la première correction non-Gaussienne qui domine la partie impaire du signal et en utilisant les prédictions de la théorie perturbative, nous obtenons des contraintes sur σ qui sont montrées sur la droite de la figure B.10 et s'écrivent $\hat{\sigma} = 0.22 \pm 0.08$ en accord avec la vraie dispersion. La précision sur la mesure de β et $D(z)$ par le biais de σ_0 peut naïvement être extrapolée pour un relevé de type Euclid qui serait autour de 0.3% sur β et 1.5% sur $D(z)$ à redshift zéro.

Limitations et perspectives Dans l'expérience proposée précédemment, il faut noter qu'aucune prise en compte des effets de masque, évolution du rapport signal sur bruit avec le redshift ou volume fini n'a été étudiée ni aucune comparaison avec les sondes déjà existantes. Une des principales limitations de ce travail a été de supposer exacte la théorie perturbative standard alors que l'on sait qu'elle est peu performante en espace des redshifts. Des améliorations sur ce point sont donc cruciales pour le futur. Une liste non-exhaustive des développements à étudier pourrait être

- améliorer la théorie perturbative en espace des redshifts à l'aide par exemple d'un modèle de type "streaming model" (see for instance [Scoccimarro, 2004, Taruya et al., 2010]) et tester le domaine de validité du modèle récent proposé par [Gil-Marín et al., 2014] dans les études basées sur la topologie;
- améliorer la prescription des plans parallèles et développer une théorie plein ciel;
- étendre la théorie à la statistique du squelette comprenant filaments et murs;
- propager les contraintes obtenues sur une estimation des paramètres cosmologiques comme l'équation d'état de l'énergie noire;
- explorer des alternatives à l'utilisation d'une série de Gram-Charlier;
- faire la statistique des erreurs;
- mettre en oeuvre la théorie sur de fausses observations réalistes et démontrer les avantages d'une sonde topologique par rapport aux autres sondes existantes (cisaillement gravitationnel, supernovae, etc);
- appliquer la théorie au cas 2D du champ de cisaillement gravitationnel.

B.3.2 Comptages de galaxies dans des cellules concentriques

Si les estimateurs topologiques mentionnés dans la section précédente présentent une robustesse intéressante, ils n'en restent pas moins difficiles à prédire au-delà du regime faiblement non-linéaire (que nous appellerons ici régime quasi-linéaire). Il n'existe qu'une observable connue qui permette de faire des prédictions analytiques dans ce régime à partir des premiers principes: les comptages de galaxies dans des cellules concentriques qui mettent en oeuvre la théorie perturbative dans une configuration hautement symétrique. En effet, cette observable, qui consiste à compter des galaxies dans des sphères concentriques et à étudier leur statistique jointe, respecte la symétrie sphérique, seule configuration dans laquelle une solution complètement non-linéaire des équations de la gravitation existe (dite solution de l'effondrement sphérique). Cette symétrie permet donc, par le biais de l'effondrement sphérique, de prédire avec une précision inattendue la statistique de comptage dans des cellules concentriques dans le régime quasi-linéaire, bien au-delà de ce qui est atteignable traditionnellement avec la théorie perturbative! Nous nous proposons de montrer ici le formalisme permettant d'obtenir les prédictions analytiques de ces observables de comptage.

Définitions Soient n sphères concentriques de rayon R_i contenant une densité de galaxies $\hat{\rho}_i$. Nous cherchons ici à étudier la statistique jointe de ces densités concentriques $\mathcal{P}(\hat{\rho}_1, \ldots, \hat{\rho}_n)$. La clé pour prédire $\mathcal{P}(\hat{\rho}_1, \ldots, \hat{\rho}_n)$ à partir des premiers principes est d'utiliser la fonction génératrice

des moments

$$\mathcal{M}_{R_1\dots R_n}(\lambda_1,\dots,\lambda_n) = \sum_{p_1,\dots,p_n=0}^{\infty} \langle \hat{\rho}_1^{p_1}\dots\hat{\rho}_2^{p_2}\rangle \frac{\lambda_1^{p_1}\dots\lambda_n^{p_n}}{p_1!\dots p_n!}\,, \tag{B.7}$$

$$= \left\langle \exp\left(\sum_i \lambda_i\hat{\rho}_i\right)\right\rangle\,, \tag{B.8}$$

qui est trivialement reliée à la fonction génératrice des *cumulants*, $\varphi_{R_1\dots R_n}(\lambda_1,\dots,\lambda_n)$, par

$$\mathcal{M}_{R_1\dots R_n}(\lambda_1,\dots,\lambda_n) = \exp\left[\varphi_{R_1\dots R_n}(\lambda_1,\dots,\lambda_n)\right], \tag{B.9}$$

de telle façon que

$$\exp\left[\varphi_{R_1\dots R_n}(\lambda_1,\dots,\lambda_n)\right] = \int \mathrm{d}\hat{\rho}_1\dots\mathrm{d}\hat{\rho}_n\,\mathcal{P}(\hat{\rho}_1,\dots,\hat{\rho}_n)\exp\left(\sum_i \lambda_i\hat{\rho}_i\right). \tag{B.10}$$

La PDF jointe des densités concentriques peut aisément être retrouvée par le biais de l'équation B.10 en faisant une transformée de Laplace inverse de la fonction génératrice des cumulants

$$\mathcal{P}(\hat{\rho}_1,\dots,\hat{\rho}_n) = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}\frac{\mathrm{d}\lambda_1}{2\pi\mathrm{i}}\dots\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}\frac{\mathrm{d}\lambda_n}{2\pi\mathrm{i}}\exp\left(-\sum_i\hat{\rho}_i\lambda_i + \varphi_{R_1\dots R_n}(\lambda_1,\dots,\lambda_n)\right). \tag{B.11}$$

La fonction génératrice des cumulants Afin de prédire la fonction génératrice des cumulants, on est amené à considérer sa *transformée de Legendre* dans le régime quasi-linéaire

$$\Psi_{R_1\dots R_n}(\rho_1,\dots,\rho_n) = \sum_i \lambda_i\rho_i - \varphi_{R_1\dots R_n}(\lambda_1,\dots,\lambda_n), \tag{B.12}$$

où les densités ρ_i sont déterminées implicitement par la condition de stationnarité

$$\lambda_i = \frac{\partial}{\partial\rho_i}\Psi_{R_1\dots R_n}(\rho_1,\dots,\rho_n)\,,\quad i=1,\dots,n\,. \tag{B.13}$$

Dans la limite de variance nulle, cette transformée de Legendre prise à deux temps différents, $\Psi(\rho_1,\dots,\rho_n;\eta)$ et $\Psi'(\rho_1,\dots,\rho_n;\eta')$, prend la *même* valeur

$$\Psi_{R_1\dots R_n}(\rho_1,\dots,\rho_n;\eta) = \Psi_{R'_1\dots R'_n}(\rho'_1,\dots,\rho'_n;\eta')\,, \tag{B.14}$$

si la masse est préservée $\rho_i R_i^3 = \rho'_i {R'_i}^3$, et si ρ'_i et ρ_i sont reliés par la dynamique de l'effondrement sphérique. En effet, on peut toujours exprimer une moyenne d'ensemble en terme des propriétés statistiques du champ de densité initial de telle sorte que

$$\exp\left[\varphi\right] = \int \mathcal{D}\tau_1\dots\mathcal{D}\tau_n\,\mathcal{P}(\tau_1,\dots,\tau_n)\exp\left(\sum_i\lambda_i\rho_i(\tau_i)\right). \tag{B.15}$$

Comme les densités courantes ρ_i peuvent provenir de contrastes initiaux différents (voir la figure 5.18), l'intégration B.15 est donc une intégrale de chemin (sur tous les chemins possibles allant des configurations initiales aux densités finales considérées) de mesure $\mathcal{D}\tau_1\dots\mathcal{D}\tau_n$ et de statistique initiale connue $\mathcal{P}(\tau_1,\dots,\tau_n)$. Si les conditions initiales sont Gaussiennes, on obtient

$$\mathcal{P}(\tau_1,\dots,\tau_n)\mathrm{d}\tau_1\dots\mathrm{d}\tau_n = \frac{\sqrt{\det\Xi}\exp\left[-\Psi(\tau_1,\dots,\tau_n)\right]}{2\pi}\mathrm{d}\tau_1\dots\mathrm{d}\tau_n\,, \tag{B.16}$$

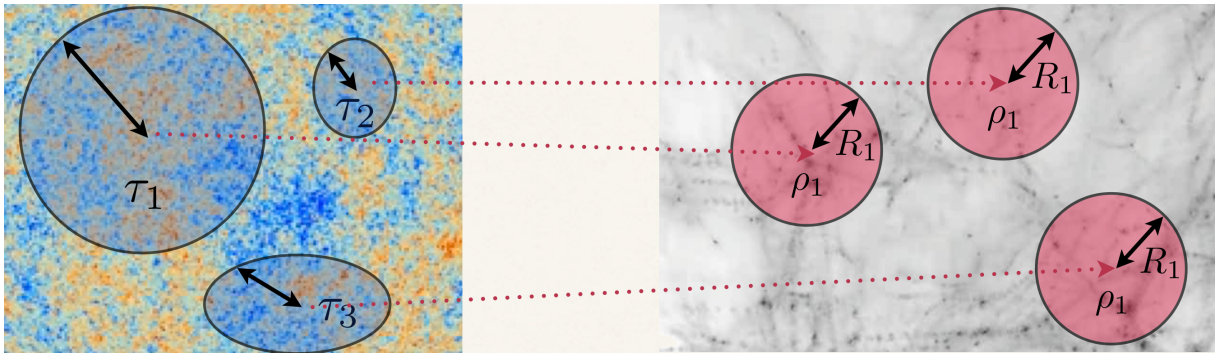


Figure B.11: Une même densité à redshift zéro (à droite) peut provenir de différentes configurations initiales (à gauche).

avec Ψ qui est donc une forme quadratique.

Dans la limite de variance nulle, l'équation B.15 est dominée par le chemin le plus probable. Comme l'observable est à symétrie sphérique, on peut s'attendre à ce que ce chemin de plus haute probabilité respecte également la symétrie sphérique et soit donc l'effondrement sphérique. Dans ce régime, l'équation B.15 devient

$$\exp\left[\varphi\right] \simeq \int \mathrm{d}\tau_1 \dots \mathrm{d}\tau_n \, \mathcal{P}(\tau_1, \dots, \tau_n) \exp\left(\sum_i \lambda_i \zeta_{\rm SC}(\tau_i)\right), \tag{B.17}$$

où le chemin le plus probable, $\rho_i = \zeta_{\rm SC}(\eta, \tau_i)$ est l'effondrement sphérique qui attribue à chaque densité finale une et une seule densité initiale. L'intégration du membre de droite de l'équation B.17 peut maintenant être mise en oeuvre par une méthode du col en approchant l'intégrale par sa valeur la plus probable où $\sum_i \lambda_i \rho_i(\tau_i) - \Psi(\tau_1, \dots, \tau_n)$ est stationnaire. Ceci nous amène à la relation fondamentale B.14 qui une fois évaluée à un temps initial arbitraire, η', donne une relation entre $\Psi(\rho_1, \dots, \rho_n; \eta)$ et les propriétés statistiques des fluctuations initiales. Lorsque celles-ci sont Gaussiennes, $\Psi(\rho_1, \dots, \rho_n; \eta_i)$ peut facilement être calculée et exprimée en terme d'éléments des matrices de variances-covariances

$$\Psi_{R_1 \dots R_n}(\rho_1, \dots, \rho_n; \eta_i) = \frac{1}{2} \sum_{i,j \leq 2} \Xi_{ij} (\rho_i - 1)(\rho_j - 1), \tag{B.18}$$

où Ξ_{ij} est l'inverse de la matrice de variances-covariances, $\Sigma_{ij} = \langle \tau_i \tau_j \rangle$, entre les contrastes initiaux $\tau_i = \rho_i - 1$ des n sphères concentriques de rayon R_i. On peut ensuite écrire la fonction génératrice des cumulants à temps quelconque en utilisant l'effondrement sphérique entre une densité finale à temps η dans une sphère de rayon R_i et un contraste initial dans une sphère centrée sur le même point et de rayon $R_i' = R_i \rho_i^{1/3}$ (par conservation de masse); formellement on peut l'écrire

$$\rho_i = \zeta_{\rm SC}(\eta, \tau_i) \approx \frac{1}{(1 - D_+(\eta)\tau/\nu)^\nu}, \tag{B.19}$$

où on utilise ici une prescription simple avec $D_+(\eta)$ le facteur de croissance linéaire et $\nu = 21/13$ pour reproduire le coefficient de dissymétrie (skewness) à haut redshift.

La PDF jointe En principe, le formalisme que nous venons de décrire permet de prédire la PDF des densités concentriques dans le régime quasi-linéaire. En réalité, seul $\Psi_{R_1, \dots, R_n}(\rho_1, \dots, \rho_n)$ est

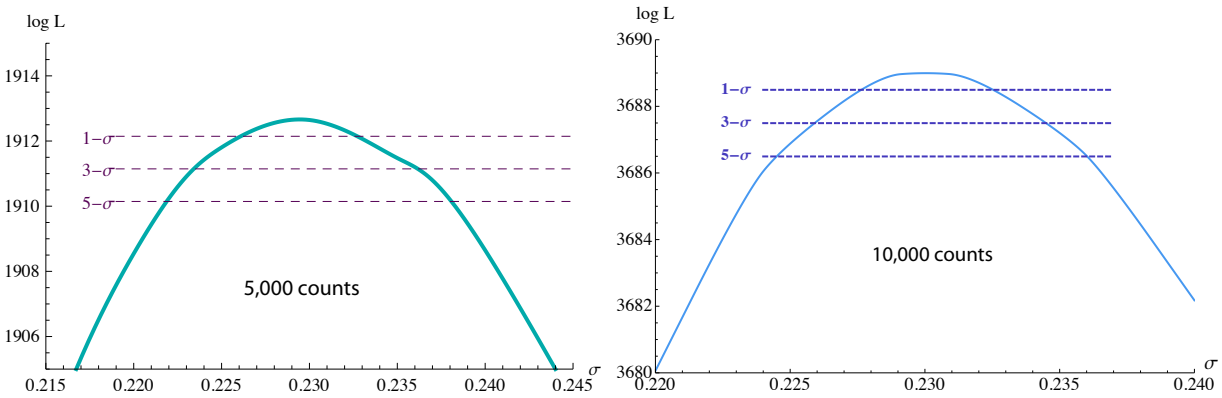


Figure B.12: Vraisemblance logarithmique d'une expérience comprenant 5000 (à gauche) ou 10 000 (à droite) sphères concentriques de rayon 10 et 11 Mpc/h mesurées dans notre simulation. Le modèle dépend ici seulement de la variance σ^2 (ν – qui paramétrise l'effondrement sphérique – et n – l'indice spectral – sont fixés). Les contours à 1,3 et 5 sigmas centrés sur la vraie valeur 0.23 sont représentés avec des tirets bleu foncé. Une expérience analogue mais en variant les trois paramètres pourrait être similairement envisagée.

facilement calculable. La fonction génératrice des cumulants, $\varphi_{R_1 \dots R_n}(\lambda_1, \dots, \lambda_n)$, est ensuite accessible via l'équation B.12 par une transformée de Legendre qui amène ses propres complications. En particulier, toutes les valeurs de λ_i ne sont pas accessibles car la relation ρ_i – λ_i ne peut pas toujours être inversée. La PDF est ensuite accessible par une transformée inverse de Laplace qui peut être calculée analytiquement dans certains régimes ou évaluée numériquement. Notons notamment qu'elle repose sur une continuation analytique dans le plan complexe de la fonction génératrice des cumulants prédite et requiert donc une connaissance fine des propriétés analytiques de $\varphi(\lambda)$. Son évaluation numérique apparaît ainsi difficile, et ce d'autant plus que le nombre de cellules considérées augmente. Ce n'est que récemment que nous avons pu évaluer avec succés la PDF à deux cellules ([Bernardeau et al., 2015]). Ceci nous a permis en particulier de prédire avec une précision surprenante les profils de densité des régions sous-denses ou sur-denses.

En fin de compte, il apparaît que la PDF des densités concentriques est sensible à la cosmologie à travers deux ingrédients: l'effondrement sphérique et le spectre de puissance linéaire, $P_k^{\rm lin}$, (via la matrice des variances-covariances, $\Sigma_{ij} = \int P_k^{\rm lin}(k) W(R_i k) W(R_j k) d^3k/(2\pi)^3$). Elle peut donc être utilisée pour tester la gravité et les modèles d'énergie noire et/ou mettre des contraintes sur les paramètres cosmologiques. Pour illustrer cette remarque, nous avons mis en oeuvre une expérience pour contraindre l'énergie noire basée sur des comptages réalisés sur des simulations ΛCDM (voir figure B.12). Avec 10 000 comptages dans deux sphères concentriques de rayon 10 et 11 Mpc/h, l'expérience imite la précision que l'on pourrait espérer dans un relevé de volume $\approx (350 h^{-1}{\rm Mpc})^3$ qui est ainsi estimée autour du pourcent.

Perspectives En résumé, la statistique de comptages dans des cellules concentriques est une sonde cosmologique prometteuse pour les prochains grands relevés de galaxies comme Euclid. L'accord entre prédictions analytiques et simulations est excellent même dans le régime quasi-linéaire où la théorie perturbative n'est traditionnellement plus valide. Une question ouverte à ce stade serait d'estimer le nombre de sphères optimal pour contraindre un ensemble donné de paramètres mais la réponse dépend probablement des caractéristiques du relevé considéré. Il faut aussi noter qu'une mise en oeuvre effective à 3D semble difficile dans la mesure où les

catalogues de galaxies en espace des redshifts brisent la symétrie sphérique locale de manière complexe. Une façon d'éviter le problème serait d'utiliser les comptages sur des observables telles que les densités projetées le long de la ligne de visée par le biais des observations de cisaillement gravitationnel ou les relevés de galaxies de type photométrique qui ne reconstruisent pas toute la statistique 3D mais uniquement la statistique circulaire 2D pour laquelle la théorie pourrait être appliquée. La précision de la méthode dans ce contexte reste à évaluer. Une autre piste à explorer pourrait être les corrélations à grande distance entre densités concentriques (et profils) qui sont accessibles par ce formalisme ([Bernardeau, 1995]) et permettraient d'avoir une théorie complète (avec en particulier une théorie des erreurs) qui pourrait être ensuite exploitée sur de vraies données.

B.3.3 Conclusion

Dans le contexte d'une cosmologie de précision, il est important de développer de nouvelles observables qui i) peuvent être prédites à partir des premiers principes; ii) sont robustes (bruit, biais,...); iii) peuvent sonder le régime quasi-linéaire de formation des structures. Nous avons mis en avant deux telles observables: d'une part les estimateurs topologiques et géométriques tels que fonctionnelles de Minkowski ou statistique des pics et d'autre part la statistique de comptage de galaxies dans des cellules concentriques qui permet d'avoir accès par exemple à la PDF de la densité cosmique ou encore aux profils de vide. Ces idées sont issues de trois articles publiés ([Codis et al., 2013, Bernardeau et al., 2014, Bernardeau et al., 2015]).

B.4 Conclusion

Dans cette thèse, nous nous sommes intéressés à l'étude théorique des grandes structures d'un point de vue à la fois analytique et numérique, la clé de voûte de ces travaux résidant dans la recherche d'une réponse aux défis posés par les futurs grands relevés de galaxies tels que la dynamique non-linéaire de formation des structures, la distorsion en espace des redshifts ou encore les alignements intrinsèques. L'ensemble de ces travaux a fait l'objet de 10 publications [Taruya et al., 2012, Codis et al., 2012, Codis et al., 2013, Laigle et al., 2015, Dubois et al., 2014, Bernardeau et al., 2014, Bernardeau et al., 2015, Codis et al., 2015a, Codis et al., 2015b, Chisari et al., 2015], d'un article qui sera soumis prochainement (Baldauf, Codis, Desjacques, Pichon, *to be submitted*) et de plusieurs articles en préparation.